THE PRESENT STATUS OF THE THEORY OF ELECTROLYTES*

CHARLES A. KRAUS

1. Introduction. It is not generally recognized that physical research is quite as dependent upon aids to the mind as it is upon aids to the senses. Indeed, it is the aids to the mind that are fashioned with greatest difficulty; only at long intervals does there appear a figure possessing that high degree of originality necessary for the construction of a new mental device that may be used successfully in analyzing the complex phenomena of nature. In chemistry, we have three such devices, namely: thermodynamics, statistical mechanics, and quantum mechanics. For the first two of these we are largely indebted to Josiah Willard Gibbs. Although a mathematician by training, Gibbs possessed that intuitive physical sense which enabled him to reduce physical phenomena to their lowest terms. His contributions to chemical science rank in importance with those of Lavoisier and Dalton.

To the layman it must appear that the rapid progress of modern science is due to experimentation. Nothing could be farther from the truth. In all ages, there have been men who experimented, just as there have been those who philosophized; but their contributions to scientific knowledge were few in number and limited in scope. What distinguishes our present age from all others is that experiment and theory have been joined in complementary fashion. The achievements of science have been due to this happy combination.

The subject that I propose to discuss is a small segment of a much broader one, namely, the relation between electricity and matter. While the problem of electrolytes is but a fraction of the broader problem, its study has led to the development of concepts which have made it possible to interpret matter in all its forms in terms of electrically charged particles.†

Interesting and instructive as it might be, an historical introduction is not possible on this occasion. Suffice it to say that the problem

---

* The fourteenth Josiah Willard Gibbs Lecture, delivered at Indianapolis, Indiana, December 28, 1937, under the auspices of the American Mathematical Society, at a joint meeting of the Society with the American Association for the Advancement of Science.

† This excepts the neutron, the relation of which to other types of ultimate particles of matter is as yet little understood.

of electrolytes had its inception in the observations of chemists, several hundred years ago, and that on the physical side it received its first impetus through the discoveries of Volta. These were followed by the epoch-making investigations of Faraday, who laid the foundations of the theory of electrolytes. He showed that matter and electricity are related in a definite, quantitative manner, whatever the nature of the atoms concerned. Most important, he introduced the concept of an ion as a charged atom or molecule. He accounted for the passage of electricity through a solution of an electrolyte by the motion of oppositely charged ions through the solution.

The ionic concept is one of the most important that has ever been introduced into physical science. It is more fundamental than the concept of atoms and molecules, for we now know that all atoms are composed of positively charged nuclei, surrounded by negative electrons equal in number to the charge on the nucleus. In final analysis, all matter is constructed of assemblies of ions. Today, the structure of the nucleus itself is interpreted in terms of ionic phenomena.

The application of the ionic concept to the problem of the structure of matter is a by-product of the problem of electrolytes. This latter problem is still far from solution in complete form. Through the work of Clausius, Hittorf, Kohlrausch, Arrhenius, and others, it finally became clear that, although electrolytes contain oppositely charged ions, these electrolytes may dissociate into their constituent ions which move about freely in the solvent medium. Arrhenius assumed an equilibrium between the free ions and the undissociated molecules of an electrolyte and sought to account for the observed properties of such a solution through displacement of this equilibrium. While his theory was successful in accounting for solutions of weak electrolytes, it led to contradictions in the case of strong electrolytes. These contradictions were not resolved until Debye introduced his theory of ion interaction some fifteen years ago.

2. **Ion interaction: thermodynamic properties.** The properties of electrolytic solutions fall into two classes: (1) those of reversible and (2) those of irreversible processes. The first class includes all thermodynamic properties such as freezing point, boiling point, vapor pressure, and the various thermodynamic energy functions; the second includes all time-dependent phenomena, such as conductance, diffusion, and transference.

According to Gibbs, equilibrium in a multiphase system requires that the thermodynamic potential of any component shall have the same value in the several phases. In systems in which reaction occurs and equilibrium exists among the reactants, the sum of the thermo-

dynamic potentials of the reactants must be equal to zero. It becomes necessary, therefore, to determine how the thermodynamic potential of an electrolyte depends upon the various thermodynamic variables.

In a dilute solution of a substance between whose molecules no forces act (an ideal solute), the thermodynamic potential of the solute varies as the logarithm of the concentration; thus

$$\mu_i = RT \log C + I,$$

where μ_i is the thermodynamic potential of the ideal solute, R the gas constant, T the absolute temperature, C the concentration in moles per liter, and I an integration constant. If forces act between the molecules of the solute, then the observed thermodynamic potential becomes

$$\mu_0 = RT \log C + RT \log f + I,$$

where $RT \log f$ is the energy due to the interacting forces between the molecules. It is evident that this energy is $RT \log f = \mu_0 - \mu_i$.

If we have an electrolyte AB which is in equilibrium with its ions A^+ and B^- according to the reaction equation $A^+ + B^- \rightleftarrows AB$, then, according to the theory of Arrhenius, we should have the equilibrium equation

$$\frac{C_{A^+} \times C_{B^-}}{C_{AB}} = K,$$

where K is the equilibrium constant. This equation is based on the assumption that no forces act between the ions, between the ions and the undissociated molecules AB, or between the molecules AB themselves. In the light of our preceding discussion, this assumption is not permissible, since the Coulomb forces, which vary as the inverse square of the distance, are constantly acting between the ions. The difficulty may be resolved by taking into account the energy due to the interaction between the ions. For a binary electrolyte, this energy term is written in the form $RT \log f^2$, in which case our equilibrium equation assumes the simple form

$$\frac{C_{A^+} \times C_{B^-}}{C_{AB}} f^2 = K.$$

If γ is the fraction of the electrolyte dissociated into ions, then $C_{A^+} = C_{B^-} = C\gamma$, and $C_{AB} = C(1-\gamma)$. Hence, our equation takes the form

$$\frac{C\gamma^2}{1-\gamma} f^2 = K. \tag{1}$$

If we know the value of f (the activity coefficient), it becomes possible to calculate the equilibrium constant K, provided γ is known.

Let us first consider the case of a completely dissociated electrolyte, ($\gamma = 1$), at relatively low concentration; the positive and negative ions will be distributed throughout the solvent medium in a somewhat random manner due to their energy of thermal agitation. This distribution of the ions, however, will not be the same as that of uncharged molecules between which no forces act. Due to the Coulomb forces acting between the ions, there is a greater probability that a positive charge will be found in the neighborhood of a negative charge than in the neighborhood of another positive charge, and the same is true of a negative charge in the neighborhood of a positive charge.

Debye has succeeded in computing the charge distribution in the neighborhood of an ion. A detailed account of the method employed would be out of place here, but I may briefly present the physical picture and give the final result. Let us fix our thoughts for the moment on a positive ion bearing a unit charge; it will, on the average (with respect to time) be surrounded by one unit of negative charge, distributed symmetrically with respect to the central ion, and the amount of charge in a shell of thickness dr, around the positive charge, will vary in a regular manner with the distance from the central charge. Actually, the charge in a shell of given thickness gradually increases from very low values, passes through a maximum, and then approaches zero at great distances. According to the Debye theory, the charge density falls off to $1/\epsilon$th its value at a distance $1/\kappa = 1.99 \times 10^{-10}(DT/C)^{1/2}$, where D is the dielectric constant of the medium. As the concentration diminishes, the distance $1/\kappa$ increases, and it follows that, in diluting a solution, work must be done against the forces acting between the central ion and its surrounding charge, or its ion atmosphere, as it is called.

For a relatively dilute solution, the equation for the activity coefficient of the ions takes the simple form

$$-\log_{10} f = \beta C^{1/2}, \tag{2}$$

where

$$\beta = \frac{1.81 \times 10^6}{(DT)^{3/2}}.$$

In our equations above, we have assumed that each ion carries a unit charge. Similar relations are easily obtained for the case of ions with any number of charges. For the sake of simplicity, our subsequent discussion will be restricted to uni-univalent salts, unless otherwise mentioned.

In Fig. 1 are plotted values of $-\log f$ against the square root of the (total ion) concentration for uni-univalent, uni-divalent, and uni-trivalent salts. As may be seen, at lower concentrations the points ap-

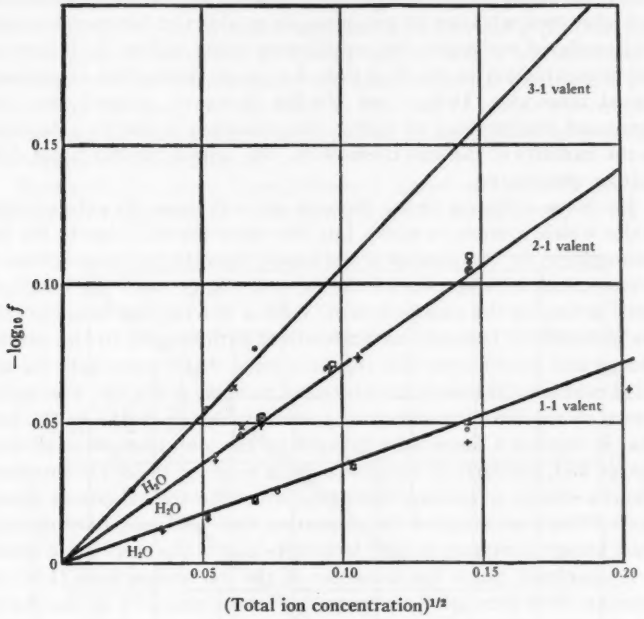


FIG. 1

Plot of $-\log f$ as a function of the square root of total ion concentration for salts in water.

proximate closely to the straight lines required by theory. As the charge on one of the ions increases, the curves become steeper in accord with theory. The equation has been tested with the very extensive experimental material relating to aqueous solutions and has been found to hold over a considerable range of concentration in the dilute region. At high concentrations, however, the deviations become large.

3. **The Debye-Hückel-Onsager equation.** Let us now turn to the irreversible phenomenon of the electrical conductance of a solution of an electrolyte. This is an important case, since the conductance is the one property of electrolytic solutions that may be measured with high precision over the widest range of conditions. It is known that

in the case of all strong electrolytes the equivalent conductance Λ (the specific conductance divided by the equivalent concentration) increases with decreasing concentration and approaches a limiting value. Arrhenius accounted for the gradually decreasing conductance at higher concentration by assuming an equilibrium between ions and undissociated molecules, the equilibrium being shifted with increasing concentration in the direction of a larger proportion of undissociated molecules. Debye and Hückel, however, showed that the decreased conductance at higher concentration is due to a decrease in the mobility of the ions themselves. The reason for this is not difficult to understand.

The force acting on an ion depends not only upon the external field under which it tends to move, but also upon the field due to the ion atmosphere. In the case of a stationary ion, the ion atmosphere is symmetrical with respect to the central charge, and the resultant force acting on the charge is zero. With a moving ion, however, the ion atmosphere becomes unsymmetrical with respect to the central charge and exerts upon this charge a force that counteracts the applied field and thus tends to reduce the mobility of the ion. The asymmetry of the ion atmosphere of a moving charge is due to the fact that it requires a finite time to build up the ion atmosphere about a charge and, similarly, it requires a finite time for the ion atmosphere about a charge to become dissipated when the central charge disappears. This time is called the relaxation time and may be computed from known constants of the electrolyte and of the medium in which it is dissolved. Since the diameter of the ion atmosphere $(1/\kappa)$ diminishes with increasing concentration, the asymmetry of the charge (in the ion atmosphere) exerts an increasingly great influence upon the moving ion as the concentration increases.

Another factor needs to be taken into account, namely, what is known as the electrophoretic effect. Due to the charge of the ion atmosphere, the water molecules in this atmosphere are carried in a direction opposite to that in which the ion moves. This results in a further diminution in the forward movement of the central charge.

The influence of the ion atmosphere on the conductance of an electrolyte, as computed by Debye and Hückel and Onsager, takes the comparatively simple form of the following equation:

$$\Lambda = \gamma(\Lambda_0 - \alpha(\gamma C)^{1/2}), \tag{3}$$

where $\alpha = 8.18 \times 10^5/(DT)^{3/2}\Lambda_0 + 82/\eta(DT)^{1/2}$. Here η is the viscosity of the solution, and Λ_0 is the limiting value which the equivalent conductance approaches as the concentration is diminished without

limit; in other words, it is the conductance of the ion when the influence of the ion atmosphere becomes vanishingly small.

In a solution of a strong electrolyte in a solvent of high dielectric constant (such as water), the electrolyte is completely dissociated, and γ, in the above equation, becomes equal to unity. In the equations originally developed by Debye and Hückel, γ was assumed to be unity, and the correctness of this assumption has been amply confirmed by the experimental facts in the case of strong electrolytes. In the case of weak electrolytes, however, it is necessary to introduce the degree of dissociation γ.

How well the Debye-Hückel-Onsager theory accounts for the conductance of solutions of electrolytes in water is illustrated in Fig. 2,

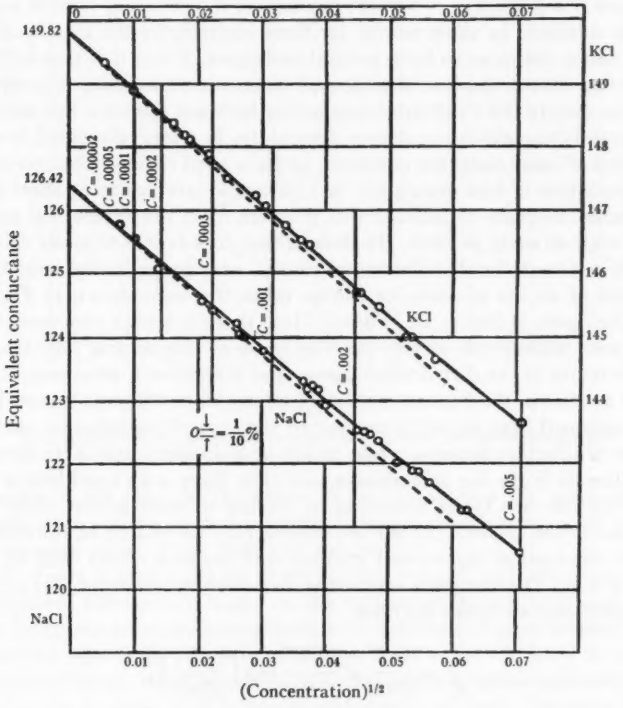


FIG. 2

Plot of equivalent conductance as a function of square root of concentration for NaCl and KCl in water at 25°C.

where the equivalent conductance of aqueous solutions of potassium chloride and sodium chloride are plotted against the square root of concentration. As may be seen from the figure, the experimental points lié upon curves that approach definite limiting tangents, shown in the figure as broken lines. These tangents, as drawn, are determined by the coefficients α and are, therefore, not arbitrary. It is evident that, while the Debye-Hückel-Onsager equation gives a correct interpretation of the facts in very dilute solutions, at higher concentrations, even as low as 0.001 N, the deviations become appreciable and become increasingly larger as the concentration increases.

4. **Ion association.** As stated above, for solutions of weak acids and bases it is necessary to assume incomplete dissociation, even in aqueous solution. In other words, for these electrolytes the ions interact at short distances to form neutral molecules. There is reason for believing that in the case of acids and bases something more is involved than merely the Coulomb forces acting between the ions. For normal electrolytes (which are strong electrolytes in water) dissolved in solvents of lower dielectric constant, we have good reason to believe that association of ions occurs due to Coulombic interaction at short distances. Bjerrum considered this problem from the theoretical point of view as early as 1926. He showed that in solvents of lower dielectric constant the distribution of ions of one charge in the neighborhood of an ion of opposite charge takes the form shown in Fig. 3. (The figure is due to Dr. Fuoss.) Here there is a high concentration of ions within one or two ion-diameters of the central ion, then a minimum in the distribution curve, and thereafter a maximum similar to that in the ion atmosphere of strong electrolytes in water. If it is assumed that an ion lying near to the central ion takes no part in the conduction process, while an ion at a distance exerts its normal influence in the ion atmosphere, and that there is an equilibrium between the two types according to the law of mass action, then the dissociation constant of the electrolyte may be related to the dielectric constant of the solvent medium and the sum of the radii of the two ions. The complete expression is somewhat complex, but a first approximation takes the form

$$\frac{1}{K} = \frac{4\pi N}{1000} a^3 \frac{aDkT}{e^2} \exp\left(\frac{e^2}{aDkT}\right),$$

where N is Avogadro's number, k is Boltzmann's constant, e is the unit charge, and a is the distance between the centers of charge of the

two ions when in contact. It is seen from this equation that the dissociation constant K is closely related to the dielectric constant D of the medium, on the one hand, and to a, the distance of closest ap-

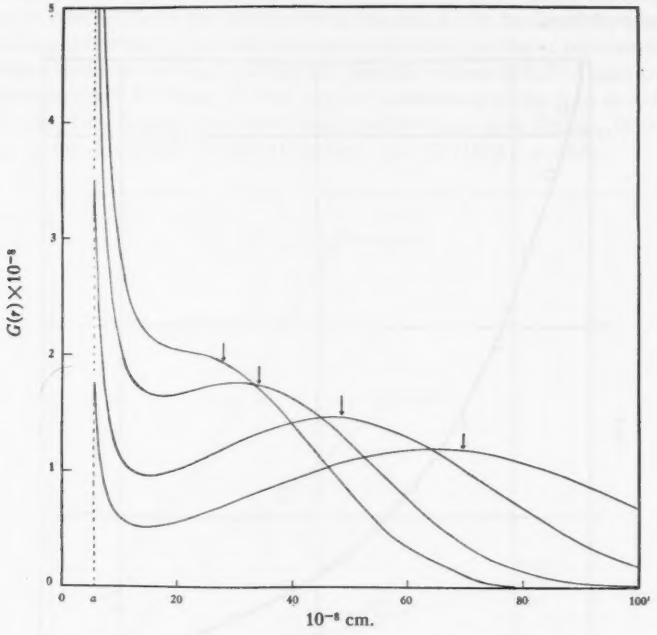


FIG. 3

Distribution of charge around central ion in solvent of dielectric constant $D=20$. $a=5.57\times10^{-8}$. $C=0.5, 1, 2, 3\times10^{-3}$.

proach of the ions, on the other. It is evident that the value of K diminishes rapidly with decreasing value of the distance a, or with decreasing value of the dielectric constant D.

In Fig. 4 is shown the curve connecting $-\log K$ with the dielectric constant. The curve is based on the value $a=6.4\times10^{-8}$ cm. and the circles represent experimental points for solutions of tetra*iso*amylammonium nitrate in mixtures of dioxane and water. As may be seen from the figure, the dissociation constant reaches values approaching unity for a value of D in the neighborhood of forty. For this and higher values of D, the electrolyte is completely dissociated. For $D=2.4$ the dissociation constant has fallen to 2×10^{-16}.

We should expect that all electrolytes in solvents of lower dielectric constant would exhibit marked association of their ions, and that the association would be the greater the smaller the ions. This has been amply confirmed by the work of Dr. Fuoss and others in the Brown laboratories.

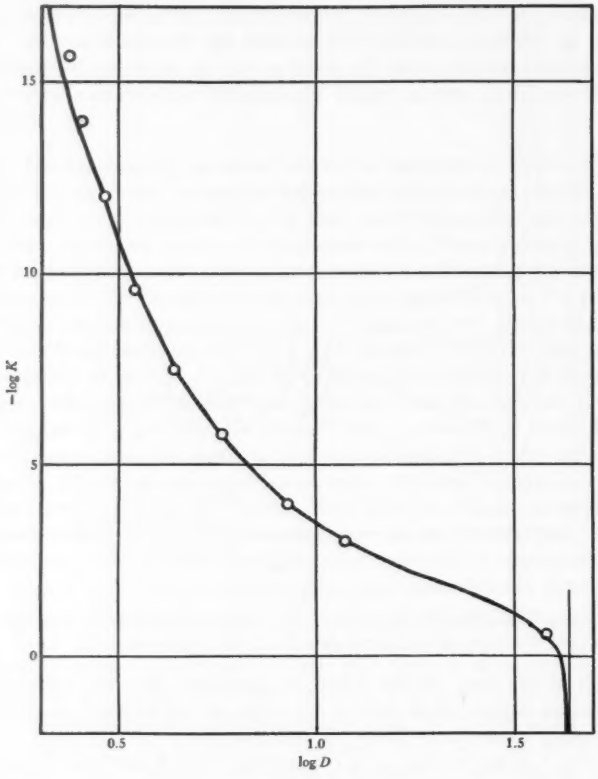


FIG. 4

Plot of $-\log K$ as function of $\log D$. $a = 6.4 \times 10^{-8}$.

In examining experimental conductance data in the light of the above theoretical considerations, it is necessary to solve simultaneously equations (1), (2), and (3) above. While an explicit solution is not possible, Dr. Fuoss has developed ready methods of approxima-

tion that render it possible to evaluate the experimental results in terms of the theory. Essentially, what is necessary, according to the present theory, is to take into account the equilibrium between free and associated ions according to the law of mass action, and the long range interaction of the ions between one another in terms of the ion-atmosphere effect. How well the theory accounts for the experimental values is shown in Fig. 5, where are plotted values of $\Lambda_0 F/\Lambda$ against those of $C\Lambda f^2/F$, where f is the activity coefficient of the ions and F is a function readily computed from conductance data (Fuoss, Journal of the American Chemical Society, vol. 57 (1935), p. 488).

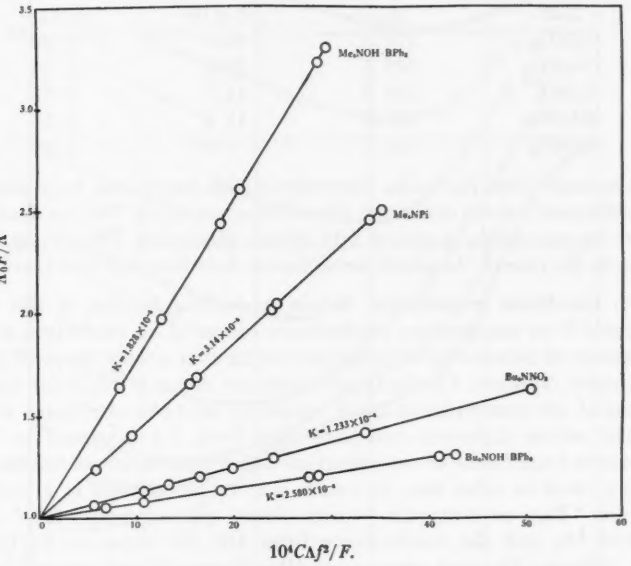


FIG. 5

Plot of conductance function for salts in ethylene chloride.

As may be seen from the figure, the experimental points closely approximate the linear relation required according to theory. At higher concentrations, however, the deviation becomes increasingly larger. Again, theory supplies a satisfactory limiting solution.

For solutions of electrolytes having spherical ions, the dissociation constant is determined by a, the sum of the ion radii, and D, the dielectric constant of the solvent medium. If K is known, a may be

computed therefrom. The values of a so obtained are in reasonable accord with atomic dimensions as determined by other means. With electrolytes having complex or unsymmetrical ions, the distance a depends upon the relative position of the ions and the simple theory (based on spherical ions) becomes an approximation. Nevertheless, the experimental values are fairly well accounted for even in this case.

In Table 1 are given values of Λ_0, K, and a for a number of electrolytes in liquid ammonia.

TABLE 1

Conductance of salts in liquid ammonia

Salt	Λ_0	$K\times10^4$	$a\times10^8$
$LiNO_3$	277	36.5	5.04
$NaNO_3$	295.1	28.8	4.46
KNO_3	329.0	14.7	3.57
$RbNO_3$	341.8	11.4	3.24
$CsNO_3$	333.5	9.66	3.21

As appears from the table, electrolytes with larger ions have lower conductance values and larger dissociation constants. The ion diameters are reasonably in accord with atomic dimension. The reversal of size in the case of the alkali metal ions is doubtless due to solvation.

5. **Ion-dipole interactions.** Before proceeding further, it will be helpful if we examine the conductance curves of an electrolyte as a function of concentration in solvent media over a wide range of the dielectric constant. Conductance curves are shown in Fig. 6 for solutions of tetra*iso*amylammonium nitrate in mixtures of dioxane and water whose dialectric constants vary from 2.2 (dioxane) to 78 (water). Logarithms of concentration and of equivalent conductance are plotted in order that all values may be represented in a single figure. The concentration varies from approximately $0.1N$ to $1\times10^{-5}N$, and the conductance from 100 for water to 3×10^{-5} for dioxane. The first two curves ($D=78$ and 37) are normal, approximating the Debye-Hückel-Onsager equation. The third curve ($D=12$) shows a large conductance decrease; here ion association is beginning to occur. With the fourth curve ($D=9$), a minimum appears in the curve at about 0.03 N. As the dielectric constant diminishes, the conductance values fall and the minimum shifts to lower concentrations, reaching 2.5×10^{-5} in pure dioxane. For dilute solutions, in mixtures having values of D less than 6, the curves approach limiting tangents with a common slope of $-1/2$. This corresponds to a mass action equilibrium between free and associated ions of a

largely associated electrolyte (γ small). The complexity of the lowermost curves at concentrations above that of the minimum point are of much interest. I shall refer to them again.

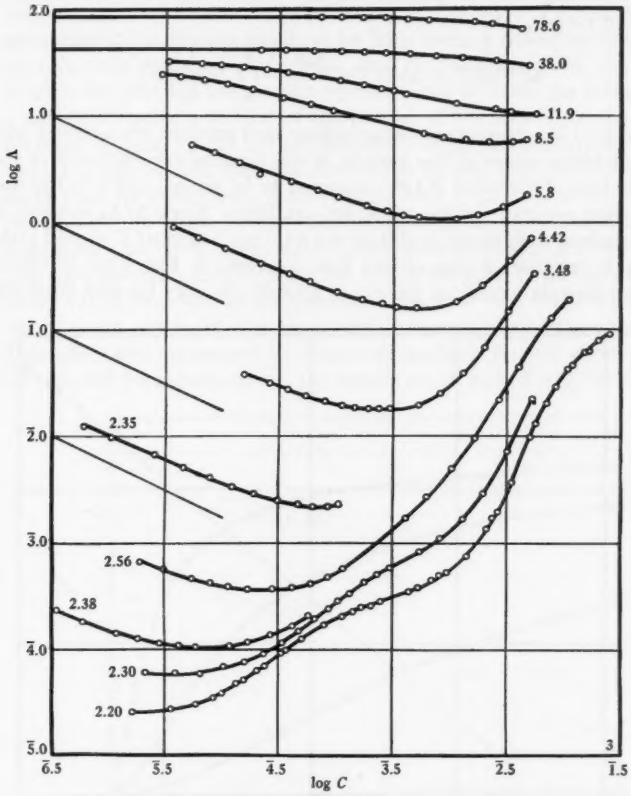


FIG. 6

Plot of log Λ as function of log C for solutions of tetra*iso*amylammonium nitrate in water-dioxane mixtures of varying dielectric constants.

The minima in the conductance curves are due to interaction between the simple ions and the associated ions or ion pairs. Thus, if we have an electrolyte AB, dissociating into the ions A^+ and B^-, the ions A^+ and B^- will interact with the ion pairs AB. An equilibrium

exists between them according to the equations $AB+A^+=ABA^+$ and $AB+B^-=BAB^-$. This interaction has been treated by Dr. Fuoss in much the same manner as that between ions and ions. It leads to the following simple equation for the conductance as a function of concentration:

$$\Lambda C^{1/2}g(C) = \Lambda_0 K^{1/2} + \frac{\lambda_0 K^{1/2}}{k}\left(1 - \frac{\Lambda}{\Lambda_0}\right)C. \tag{4}$$

Here $g(C)$ is a correction factor taking into account the activity and the mobility terms of the ions; λ_0 is the limiting conductance of the triple ions ABA^+ and BAB^-, assumed to be equal; and k is the dissociation constant of the triple ion equilibria. Since Λ/Λ_0 is small in comparison with unity, a plot of the $\Lambda C^{1/2}g(C)$ against C should yield a linear relation. A plot of this kind is shown in Fig. 7 for solutions of tetrabutylammonium picrate in anisole. As may be seen from the

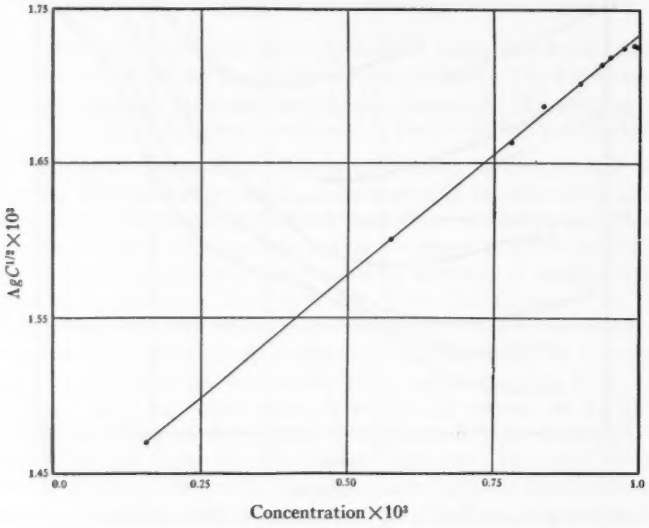


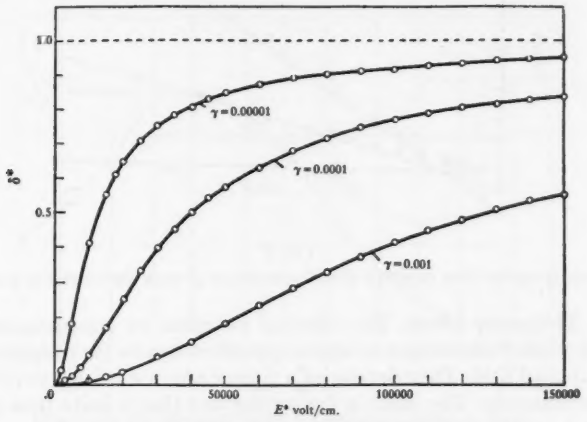
FIG. 7

Plot of conductance function for tetrabutylammonium picrate in anisole showing effect of triple-ion equilibrium.

figure, the experimental points lie upon the required straight line well within the limit of experimental error. At higher concentrations,

slightly above the concentration at the minimum point, deviations become appreciable and increase with increasing concentration.

6. **The effect of potential.** An ion moving under the action of a constant applied field is retarded owing to the asymmetry of its ion atmosphere. On the one hand, as we have seen, a potential develops opposing the potential of the field, and, on the other hand, the electrophoretic effect of the ion atmosphere tends to move the solvent in a direction opposite to that of the motion of the ion. Since the time required to build up the ion atmosphere is finite, the condition arises, when the applied field becomes sufficiently strong, that the ion moves away from its atmosphere more rapidly than it (the ion atmosphere) can be built up. In the limiting case, the ion moves quite free of an ion atmosphere and it is in the same condition as it would be if the solution were infinitely dilute. Since the relaxa ion time of the ion atmosphere varies inversely as the concentration, this field effect is greater (for a given potential) the lower the concentration of the solution. This effect was discovered by Wien and has been studied extensively by him and his collaborators; the results are in accord with theory.



γ = Concentration in moles per liter.

FIG. 8

Illustrating the Wien potential effect for solutions of strong electrolytes in water.

The effect of potential on the conductance of aqueous solutions of strong electrolytes is illustrated in Fig. 8, where applied potentials are plotted as abscissas against the increase of conductance due to the

field divided by the maximum increase possible due to an infinite field. Since the relaxation time is an inverse function of concentration, the conductance change on applying the field reaches its limiting value much more rapidly at low concentrations than at high concentrations.

In the case of weak electrolytes, another effect (due to the field) appears. Owing to the field which tends to separate the oppositely charged ions in their ion pairs, the dissociation of the electrolyte is increased as the field increases. This effect has also been studied by Wien, and the theory has been developed by Onsager.

In Fig. 9 are shown curves for the percentage conductance change of several weak acids in aqueous solution. The experimental points are indicated and the straight lines, as drawn, represent the course of the curves required by theory.

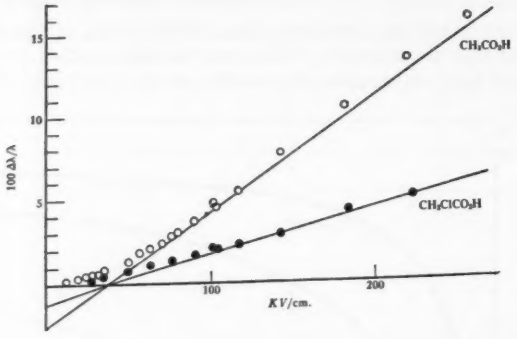


FIG. 9

Illustrating the Wien potential effect for solutions of weak electrolytes in water.

7. **Frequency effect.** The effect of potential on conductance led Debye and Falkenhagen to suspect an effect due to the frequency of the applied field. They developed a theory which was later confirmed experimentally. The effect is due to the fact that a finite time is required for the asymmetry of the ion atmosphere around a moving ion to disappear when the field is removed. With a slowly alternating field the effect is inappreciable, but when the period of the field approaches the relaxation time of the ion atmosphere, the asymmetry effect of the ion atmosphere on conductance diminishes and disappears at sufficiently high frequencies because the displacement of the central ion in its atmosphere becomes inappreciable in the period of

an oscillation. The electrophoretic effect due to the ion atmosphere itself remains.

This conductance change over that of the conductance in a stationary field decreases with decreasing concentration. It likewise decreases with decreasing number of charges on the central ion. If the conductance of an electrolyte is measured at a given frequency comparable with the relaxation time of the ions at intermediate concentrations, then, at very low concentrations where the radius of the ion atmosphere is very large, the frequency effect will be zero, since the ions will already have a conductance equal to that of their limiting value for an infinitely dilute solution. With increasing concentration, the conductance change due to frequency increases, passes through a maximum, and thereafter decreases. The reason for the decrease is that the frequency of the applied field is now below that required in order to prevent the building up of the asymmetry of the ion atmosphere.

In Fig. 10 are shown curves for the percentage change in the conductance of magnesium sulfate and manganese sulfate in aqueous solution for a field frequency corresponding to a wave length of one

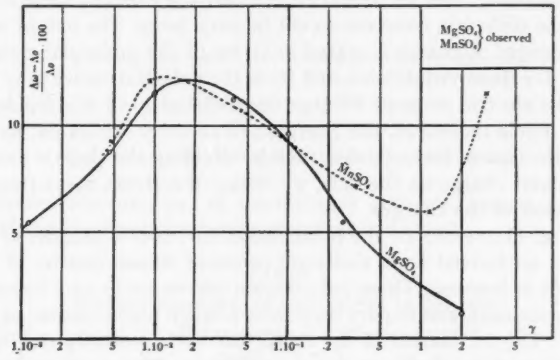


FIG. 10
Illustrating the frequency effect for solutions of salts in water.

meter. For magnesium sulfate, the curve approximates closely to the theoretical value over the entire range of concentration; for manganese sulfate, a large deviation occurs at the higher concentrations. In the case of simpler salts, such as uni-univalent salts, the correspondence between theory and experiment is even better than in that of the two salts above.

8. **Ion dipoles.** In measuring the conductance of solutions, we obtain information relating directly to the ions, but our knowledge with respect to the associated ions or ion pairs depends upon inference. It is possible, however, to carry out measurements in which the property measured depends directly upon the ion pairs and not upon the ions.

The dielectric constant of a vacuum is unity; that of all material media is greater than unity. This is due to the fact that the charges in the atoms of molecules of the medium suffer displacement under the action of an impressed potential. With non-polar molecules, this displacement of charge is extremely small, but if the centers of charge of positive and negative electricity in a molecule are not coincident (polar molecules), then the molecule has an electric moment, and it will tend to rotate under the action of the field so as to align the axis of its electrical dipole with the axis of the field. The presence of such dipole molecules in a solvent medium tends to increase the dielectric constant of the solution over that of the pure medium.

If the charges in the dipole are actually separated, as we should expect them to be in the case of associated ions, or ion dipoles, as we may call them, then we should expect that the effect of these dipoles upon the dielectric constant would be very large. The extent of the displacement of charge is stated in terms of the molecular polarization of the dissolved dipoles, and from the polarization we may compute the electric moment (charge times distance) of the dipoles. If, in the dipole molecules, the charges are actually separated, as they are in the case of ionic dipoles, then by dividing the electric moment by the unit charge on the ions, we obtain therefrom the distance of separation of the charges.

In Fig. 11 are shown the polarization curves of a number of electrolytes as derived from dielectric constant measurements of their solutions in benzene. These polarization curves are in part linear and slightly inclined, and in part they show a much higher degree of complexity. The inclination of the linear curves is probably due to the association of the dipoles to form quadrupoles. The rapid diminution of the polarization in the case of other electrolytes is doubtless due to a high degree of association of the dipoles to more complex molecules.

What is striking, however, is the fact that the values of the polarization for electrolytes dissolved in benzene are approximately ten times as great as those of ordinary polar molecules in the same solvent. These high values of the polarization indicate that the charges are separated by considerable distances. Values of the polarization, of

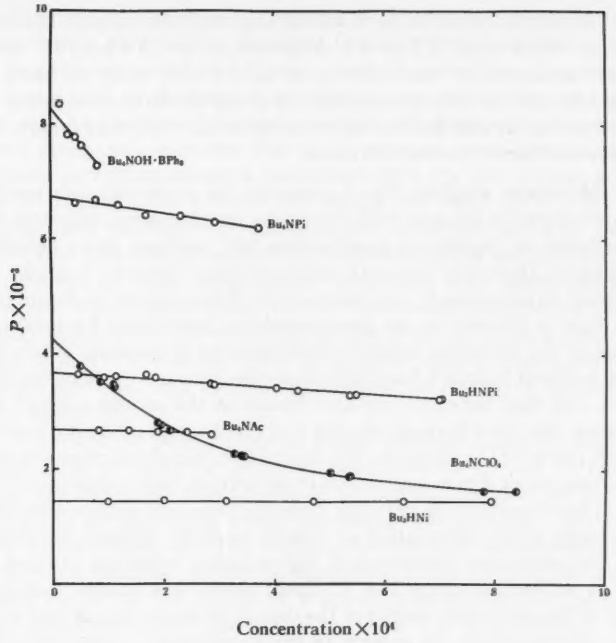


FIG. 11
Polarization curves of electrolytes in benzene solution.

the dipole moments, and of the distance between charges, as computed from the dipole moments, are given in Table 2.

TABLE 2
Electric moments of ion dipoles in benzene

Salt	*Polarization*	*Moment*$\times 10^{18}$	$a \times 10^{8}$
$Bu_4N \cdot Ac$	2690	11.2	2.34
$Bu_4N \cdot Br$	3250	12.3	2.58
$Bu_3NH \cdot Pi$	3770	13.2	2.78
$Am_3NH \cdot Pi$	3830	13.3	2.80
$Am_4N \cdot SCN$	5050	15.4	3.23
$Bu_4N \cdot Pi$	6740	17.8	3.74
$Am_4N \cdot Pi$	7090	18.3	3.83
$Am_4N \cdot HOBPh_3$	8200	19.6	4.12

Bu = butyl, Am = amyl, Ph = phenyl, Ac = acetate, Pi = picrate.

On inspection of the table, it will be seen that the distance between charges varies from 2.3 to 4.1 Ångström units (Å $=1\times10^{-8}$ cm.). These values are in reasonably good accord with what we know of molecular and atomic dimensions. It is particularly interesting to note that as the ions become more complex, and therefore larger, the distance between the ions increases.

9. **Molecular weights.** The decrease in the molecular polarization of electrolytes in benzene with increasing concentration indicates the association of dipoles to quadrupoles and perhaps more complex structures. Dielectric constant measurements, however, can give us no direct evidence as to the existence of these structures. Some further light is thrown on the phenomenon of association by measurements of the molecular weights of electrolytes in benzene. While the force between ions and ions falls off as the inverse square of the distance, and that between ions and dipoles as the inverse cube of the distance, the force between dipoles and dipoles falls off as the inverse fourth power of the distance. We should not, therefore, expect to find a high degree of dipole association in solvents of higher dielectric constant; but in solvents of very low dielectric constant, such as benzene, we should expect association of dipoles to occur. Indeed, we might look for association complexes of higher order, including charged as well as uncharged complexes. If dipole association occurs, measurement of the molecular weight of the dissolved electrolyte should indicate, approximately, its extent. In this connection, it may be noted that the concentration of ions in solutions of electrolytes in benzene is so low that the ion atmosphere effect may be neglected.

In Fig. 12 are shown curves in which are plotted values of the

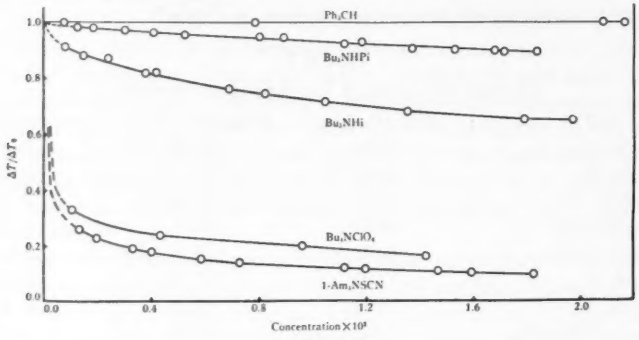


FIG. 12

Freezing point curves of solutions of electrolytes in benzene.

ratio $\Delta t/\Delta t_0$ of the observed freezing point lowering to the lowering of an ideal solute having the same (stoichiometric) concentration.

It will be observed that in the case of all electrolytes the curves lie below that of an ideal solute, indicating a measurable amount of association. An analysis of the curve for tributyl-ammonium picrate (the least associated) indicates that there exists an equilibrium between dipoles and quadrupoles in accordance with the law of mass action. In the case of other electrolytes, association takes place to a larger extent, and the observed values cannot be accounted for on the assumption of a simple equilibrium between dipoles and quadrupoles. Evidently, in these cases, higher complexes than quadrupoles are formed. This is particularly true of tetra*iso*amylammonium thiocyanate, where the mean molecular weight of the solute at a concentration of 0.017 N is eleven times its formula weight. There can be no doubt but that very complex aggregates are formed in solutions of this electrolyte.

The results of these molecular weight determinations are in accord with those of dielectric constant measurements. It will be recalled that the molecular polarization of tetra*iso*amylammonium thiocyanate falls off very sharply with increasing concentration. The results are also in accord with the results of conductance measurements. In Fig. 6, for example, for tetra*iso*amylammonium nitrate, the conductance curve for low values of D shows several inflection points at concentrations above the minimum; there is a certain region where the conductance increases very slowly with concentration. A similar curve in the case of tetra*iso*amylammonium thiocyanate shows an even more marked effect; there is a considerable region where there is scarcely any increase in conductance as the concentration is increased. Presumably, what is happening here is the formation of a large proportion of highly complex neutral structures which, while taking no part in conductance, tend to reduce the normal increase of ions with increasing concentration.

Since in all cases the conductance ultimately rises very rapidly (at high concentrations), it follows that in these regions there must be a rapid increase in the number of ions. Seemingly, there is an equilibrium among a great variety of complex ions and neutral molecules which, with increasing concentration, is shifted in the direction of an increased number of charged structures, that is, structures in which there is a preponderance of either positive or negative ions.

10. **Conclusion.** The preceding discussion will have shown that the theory based on the assumption of Coulombic interaction between

ions gives an uncommonly satisfactory account of the properties of electrolytic solutions in the limiting case of dilute solutions. It must not be forgotten, however, that the theory is strictly a limiting theory, and that as yet we have no adequate theory of more concentrated solutions. While a large mass of data is available with respect to concentrated solutions, an adequate theory is lacking. That the present theory may be expanded in some simple fashion so as to account for the more concentrated solutions seems doubtful. It appears more probable that an entirely new method of attack will have to be devised.

Actually, only a beginning has been made in the solution of the problem of electrolytes. The nature of the ions themselves remains unknown. We know that ions in water are hydrated, but the extent of the hydration is not known. There is evidence indicating that ions in media other than water are solvated, but again the extent of solvation remains unknown.

It is known that the conductance of an electrolyte in a given medium is roughly proportional to the fluidity of the medium, but the precise relation between conductance and viscosity remains undetermined. So also the relation between the conductance phenomenon and temperature remains very obscure. Indeed, with the exception of the measurements carried out by A. A. Noyes and his co-workers thirty years ago, we have practically no data with respect to aqueous solutions at temperatures above 100°, and even at that temperature the data are meagre and uncertain. In solvents other than water, there is almost a complete lack of data over any considerable temperature range.

The generality of electrolytic phenomena is still not recognized by many investigators. Practically all liquid media are capable of dissolving suitable electrolytes and yielding solutions that conduct the electric current. Much work remains to be done with solvents of very low dielectric constant. Data with respect to solutions in solvents of dielectric constant higher than that of water are also lacking.

Many substances which are not themselves electrolytes in the pure state yield electrolytic solutions when dissolved in suitable solvents. Such is the case with the acids and many bases in water. Even today we have no sound theory to account for the influence of constitution upon the strength of acids and bases. The properties of acids and bases in solvents other than water remain uninvestigated. So also the properties of the weaker salts in solvents other than water remain unknown. The investigation of the remarkable solutions of the alkali metals in liquid ammonia has been carried only far enough to show

that these solutions form a link between metallic and electrolytic conductors. These solutions offer a tempting field for further research.

Electrolytic solutions find applications in many diverse fields of science, particularly in the biological sciences. Biological systems, in general, are electrolytic systems and many biological phenomena directly involve electrolytes. It does not seem probable that biological processes may be accounted for merely on the basis of what is known of aqueous solutions. Cell structures consist of anything but water and yet these cell structures are of primary importance in connection with many electrolytic phenomena that accompany biological change. The transmission of nerve impulses is recognized to be an electrical phenomenon. Clearly, the electric impulse in this case is not a metallic one; the process must, necessarily, be an electrolytic one. An understanding of the mechanism of such phenomena will be had only when the nature of electrolytic processes in cells is more thoroughly understood.

BROWN UNIVERSITY

TORSION OF REGIONS BOUNDED BY CIRCULAR ARCS*

I. S. SOKOLNIKOFF AND E. S. SOKOLNIKOFF

1. Introduction. N. Muschelišvili discussed recently† some elegant general methods of solution of a class of problems occurring in the theory of elasticity. These methods will bear closer scrutiny as the present authors have convinced themselves by obtaining, with remarkable ease, solutions of St. Venant's torsion problem for some regions bounded by circular arcs. A number of problems in harmonic and biharmonic analysis, recently discussed by other investigators, can be treated with greater simplicity by these methods.

Muschelišvili makes use of the fact that a function of a complex variable ζ, analytic for $|\zeta|<1$, whose real part on $|\zeta|=1$ assumes continuous real values $u=u(\theta)$, $(0\leqq\theta<2\pi)$, is given by

$$f(\zeta)=\frac{1}{\pi i}\int_{\gamma}\frac{u(\theta)d\sigma}{\sigma-\zeta}+\text{const.}, \tag{1}$$

where $\sigma=e^{i\theta}$, and ζ is any point interior to the unit circle γ. It is readily shown that (1) is entirely equivalent to the formula of Schwarz.

Consider a continuous, simple closed curve C in the complex z plane and let $z=w(\zeta)$ map the region bounded by C conformally on the unit circle γ in the ζ plane. Let $F(z)$ be analytic in the interior of C, and let the value of the real part of $F(z)$ on the boundary of C be $u(x, y)$.

Substituting

$$x=\frac{w(\zeta)+\bar{w}(\bar{\zeta})}{2}, \qquad y=\frac{w(\zeta)-\bar{w}(\bar{\zeta})}{2i}$$

in $u(x, y)$ gives

$$u(x, y)=\phi(\sigma, \bar{\sigma})=\phi\left(\sigma, \frac{1}{\sigma}\right),$$

where bars denote the conjugate values. Hence, from (1),

---

* Presented to the Society, December 30, 1937.

† N. Muschelišvili, Rendiconti delle Reale Accademia dei Lincei, vol. 9 (1929), pp. 295–300; Mathematische Annalen, vol. 107 (1932), pp. 282–312; Zeitschrift für Angewandte Mathematik und Mechanik, vol. 13 (1933), pp. 264-282.

$$(2) \qquad F(z) = F[w(\zeta)] = \frac{1}{\pi i}\int_{\gamma} \frac{\phi\left(\sigma, \frac{1}{\sigma}\right) d\sigma}{\sigma - \zeta} + \text{const.},$$

and the problem of Dirichlet for the region bounded by C is solved.

It may be remarked that the procedure just outlined differs from that formerly used by one of the present authors* in that the integral of Poisson employed by him is replaced by a modified integral of Schwarz. The latter frequently admits of easy evaluation, in closed form, as is illustrated below in the treatment of a torsion problem.

2. **Torsion problem.** Consider a region R of the complex z plane bounded by two circular arcs C_1 and C_2 making an angle $\alpha \neq 0$ at their points of intersection $z = z_1$ and $z = z_2$. The function effecting conformal mapping of the region R on the unit circle in the complex ζ plane is

$$(3) \qquad \zeta = \frac{Ci(z - z_1)^{\pi/\alpha} + (z - z_2)^{\pi/\alpha}}{Ci(z - z_1)^{\pi/\alpha} - (z - z_2)^{\pi/\alpha}}.$$

The choice of the constants C, z_1, and z_2 is uniquely determined by the geometrical configuration and the scale used in mapping the region R on the unit circle. For $\alpha = \pi/n$, where n is an integer, the mapping function becomes rational, which fact greatly simplifies the evaluation of (2). To illustrate the procedure it will suffice to consider two important special cases when the region R is one of the following:

(a) a lune formed by two circular arcs of equal radius and intersecting at right angles;

(b) a semi-circle.

The first of these regions suggests the profile of an airplane wing. It is easily verified that the mapping functions appropriate to the regions defined in (a) and (b) are, respectively,

$$(4) \qquad \zeta = \frac{2z}{z^2 + 1}$$

and

$$(5) \qquad \zeta = \frac{(z + 1)^2 - i(z - 1)^2}{(z + 1)^2 + i(z - 1)^2}.$$

In obtaining (4) and (5) the constants z_1 and z_2 were chosen as -1 and $+1$, which can be done without loss of generality inasmuch as

* I. S. Sokolnikoff, Transactions of this Society, vol. 33 (1931), pp. 719–732.

this choice merely fixes the scale for the radii of arcs bounding the regions. The constant C was chosen so as to make the point $z=0$ in (4) correspond to $\zeta=0$, and the point $z=0$ in (5) correspond to $\zeta=-i$.

The transformations (4) and (5) permit of easy solution, in closed forms, of the torsion problem for prisms whose cross sections are the regions defined in (a) and (b). The solution of the second of these problems is included in the papers of Greenhill,* who attacked the problem in an entirely different way. The solution of the first problem appears to be lacking, and it will be given next.

Since the conjugate $\Psi(x, y)$ of the torsion function $\Phi(x, y)$ assumes on the boundary of the cross section of the prism the value

$$\begin{aligned}\Psi &= \tfrac{1}{2}(x^2+y^2)+\text{const.}\\ &= \tfrac{1}{2}z\bar{z}+\text{const.}=\tfrac{1}{2}w(\sigma)\bar{w}(\bar{\sigma})+\text{const.},\end{aligned}$$

it follows from the foregoing that the real part of the analytic function

$$f(\zeta)=\frac{1}{2\pi i}\int_{\gamma}\frac{w(\sigma)\bar{w}(\bar{\sigma})}{\sigma-\zeta}\,d\sigma+\text{const.} \tag{6}$$

gives the desired function $\Psi(x, y)$. From (4) it follows that $w(\sigma)=[1-(1-\sigma^2)^{1/2}]/\sigma$, where the appropriate branch of the square root is determined from the observation that the imaginary part of z is positive whenever θ in $\sigma=e^{i\theta}$ lies between 0 and π.

Noting that on $|\zeta|=1$ we have $\bar{\sigma}=1/\sigma$, we deduce the relation

$$w(\sigma)\bar{w}(\bar{\sigma})=[1-(1-\sigma^2)^{1/2}]\left[1-\left(1-\frac{1}{\sigma^2}\right)^{1/2}\right].$$

Substituting this expression in the integrand of (6) and evaluating the resulting elementary integrals, we obtain†

$$f(\zeta)=-(1-\zeta^2)^{1/2}-\frac{1-\zeta^2}{\pi\zeta}\log\frac{1-\zeta}{1+\zeta}+\text{const.},$$

or if we return to the z-plane,

$$F(z)=\frac{z^2-1}{z^2+1}-\frac{(1-z^2)^2}{\pi z(1+z^2)}\log\frac{1-z}{1+z}+\text{const.}$$

---

* A. G. Greenhill, Messenger of Mathematics, vol. 8 (1879), p. 89; vol. 10 (1881), p. 83.

† The details of integration are of little interest and hence are omitted. In evaluating the integrals involving square roots, it is advantageous to make a substitution $\sigma=(1+it^2)/(1-it^2)$, so that the resulting integrals are evaluated over the real axis of the t plane.

The real part of $F(z)$, which is the desired solution, is

$$\Psi(x, y) = \{\pi(x^2+y^2)[(x^2+y^2+1)^2-4y^2]\}^{-1}\{\pi(x^2+y^2)[(x^2+y^2)^2-1] + x(x^2+y^2+1)[2y^2-(x^2+y^2-1)^2]\log R + y(x^2+y^2-1)[(x^2+y^2+1)^2+4x^2]\phi\},$$

where

$$R = \frac{[(1-x^2-y^2)^2+4y^2]^{1/2}}{(1+x)^2+y^2} \quad \text{and} \quad \phi = \tan^{-1}\frac{-2y}{1-x^2-y^2}.$$

A simple calculation shows that, on the boundary of the lune formed by $x^2+(y\pm 1)^2 = 2$, $\Psi(x, y)$ reduces to

$$\Psi = \frac{x^2+y^2}{2} + \text{const.},$$

as it should.

An analogous calculation* gives for the region bounded by a semicircle

$$f(\zeta) = \frac{2^{1/2}(1+i\zeta)(1-\zeta^2)^{1/2}}{(i+\zeta)^2} + \frac{4}{\pi i(i+\zeta)} + \frac{2}{\pi i}\frac{(1-\zeta^2)}{(i+\zeta)^2}\log\frac{i(1-\zeta)}{1+\zeta} + \text{const.},$$

or, if we return to the z-plane with the aid of (5),

$$F(z) = \frac{1}{2\pi i}\left\{\pi i z^2 + \frac{2(z^2+1)}{z} + \frac{(z^2-1)^2}{z^2}\log\frac{1-z}{1+z}\right\} + \text{const.},$$

which agrees with the result of Greenhill in the form given by Love.†

A similar procedure can be employed in treating St. Venant's flexure problem, the solution of which, for the regions in question, does not appear in the literature.‡

THE UNIVERSITY OF WISCONSIN

* The integrals appearing in this case are precisely of the same type as those occurring in the preceding problem. They are evaluated with the aid of the elementary theorems on residues, and by means of the substitution given in the preceding footnote.

† A. E. H. Love, *A Treatise on the Mathematical Theory of Elasticity*, 4th ed., 1927, p. 319.

‡ See, however, M. Seegar and K. Pearson, Proceedings of the Royal Society, Section A, vol. 96 (1920), pp. 211–232.

A MULTIPLIER RULE IN ABSTRACT SPACES*

HERMAN H. GOLDSTINE

The Lagrange multiplier rule has been generalized to certain non-calculus of variations problems by Graves,† Hahn,‡ and the author.§ Moreover a very general problem was formulated by L. A. Lusternik.‖ However his work seems to rest upon a theorem which is stated without proof and which the author is unable to verify. There are also certain other difficulties with his proof. The problem herein presented is so formulated that the problem of Bolza in the calculus of variations, the problems treated by Graves, Hahn, and the author, and the problem of minimizing a functional defined upon an arbitrary Banach space subject to very general numerically-valued side conditions, and numerous other examples are included as special cases.

The proof proceeds along lines which are essentially generalizations of the methods of the calculus of variations. This demonstration is made possible by the very powerful implicit function theorems of Hildebrandt and Graves¶ and yields analogs of the transversality condition and of the Euler-Lagrange equations. Some instances which explain the number of linear spaces involved are given in the concluding section.

1. **Definitions and assumptions.** In order to obtain our statement and proof of the multiplier rule it is convenient to refer to five normed linear spaces $\mathfrak{M}$, $\mathfrak{N}$, $\mathfrak{X}$, $\mathfrak{U}$, $\mathfrak{V}$, of which $\mathfrak{M}$, $\mathfrak{U}$, and $\mathfrak{V}$ are assumed to be complete. It is further supposed that $\mathfrak{M}_0$, $\mathfrak{N}_0$, and $\mathfrak{X}_0$ are regions of their respective spaces. We shall then be concerned with the following basis:

* Presented to the Society, September 10, 1937.

† *A transformation of the problem of Lagrange in the calculus of variations*, Transactions of this Society, vol. 35 (1933), pp. 675–682.

‡ *Ueber die Lagrange'sche Multiplikatorenmethode*, Sitzungsberichte der Akademie Wien, vol. 131 (1922), pp. 531–550.

§ *The minima of functionals with associated side conditions*, Duke Mathematical Journal, vol. 3 (1937), pp. 418–425.

‖ *Sur les extrêmes relatifs des fonctionnelles* (in Russian), Recueil Mathématique de la Société Mathématique de Moscou, vol. 41 (1934), pp. 390–401.

¶ See L. M. Graves, *Topics in the functional calculus*, this Bulletin, vol. 31 (1935), pp. 641–662; *Riemann integration and Taylor's theorem in general analysis*; *Implicit functions and differential equations in general analysis*; and T. H. Hildebrandt and L. M. Graves, *Implicit functions and their differentials in general analysis*. The last three articles cited appear in the Transactions of this Society, vol. 29 (1927), pp. 163–177, 514–552, and 127–153, respectively. In the sequel these papers will be referred to as papers I to IV, respectively.

A. f and ψ_k, $(k=1, \cdots, q)$, where q is an arbitrary integer, are real-valued functions defined on the composite* $(\mathfrak{X}_0, \mathfrak{M}_0)$ of $\mathfrak{X}_0$ and $\mathfrak{M}_0$, each of which has a first differential at some fixed point $(\bar{x}_0, \bar{\mu}_0)$. Moreover the functionals ψ_k, $(k=1, \cdots, q)$, vanish at this point $(\bar{x}_0, \bar{\mu}_0)$.

B. $u=\phi_1(x_0, \mu_0, \nu_0)$ is an operation of class C' defined on $(\mathfrak{X}_0, \mathfrak{M}_0, \mathfrak{N}_0)$, and having its functional values in $\mathfrak{U}$; and $\nu_0=n(x_0, \mu_0)$ is defined over $(\mathfrak{X}_0, \mathfrak{M}_0)$, has its values in $\mathfrak{N}_0$, is of class C', and is such that $\phi_1[\bar{x}_0, \bar{\mu}_0, n(\bar{x}_0, \bar{\mu}_0)]=0_u$, the zero element in the space $\mathfrak{U}$. (Analogous notations will be used for the zero points in the various spaces treated.)

C. There is an operation $v=\phi_2(x_0, \mu_0, \nu_0)$ of class C', which is defined on a neighborhood of the point $(\bar{x}_0, \bar{\mu}_0, \bar{\nu}_0)$, where $\bar{\nu}_0=n(\bar{x}_0, \bar{\mu}_0)$, and has its functional values in $\mathfrak{V}$, such that the linear and continuous function of μ, $(d_\mu\phi_1(x_0, \mu_0, \nu_0; \mu), d_\mu\phi_2(x_0, \mu_0, \nu_0; \mu))$, has a linear continuous reciprocal at $(x_0, \mu_0, \nu_0)=(\bar{x}_0, \bar{\mu}_0, \bar{\nu}_0)$.

D. The functional f has a minimum at $(\bar{x}_0, \bar{\mu}_0)$ in the class of points in $(\mathfrak{X}_0, \mathfrak{M}_0)$ which satisfy the equations

$$\phi_1[x_0, \mu_0, n(x_0, \mu_0)] = 0_u,$$
$$\psi_k(x_0, \mu_0) = 0, \qquad k = 1, \cdots, q.$$

In order to simplify the notation the element $\bar{v}$ of $\mathfrak{V}$ will be used to designate $\phi_2(\bar{x}_0, \bar{\mu}_0, \bar{\nu}_0)$; $\Phi_\alpha(x, \mu, \nu)$ to indicate $d\phi_\alpha(\bar{x}_0, \bar{\mu}_0, \bar{\nu}_0; x, \mu, \nu)$, $(\alpha=1, 2)$; and $N(x, \mu)$ to connote $dn(\bar{x}_0, \bar{\mu}_0; x, \mu)$.

2. **The imbedding theorem.** It is of paramount importance for our purposes to show that the minimizing element $(\bar{x}_0, \bar{\mu}_0)$ is a limit point of elements (x_0, μ_0) which satisfy the equation

$$\phi_1[x_0, \mu_0, n(x_0, \mu_0)]=0_u.$$

To show that the element $(\bar{x}_0, \bar{\mu}_0)$ is so imbedded it is convenient first to demonstrate several lemmas.

LEMMA 2.1. *In a sufficiently small neighborhood of $\bar{\mu}_0$ the equations*

$$\phi_1(x_0, \mu_0, \nu_0) = u,$$
$$\phi_2(x_0, \mu_0, \nu_0) = v \tag{2.1}$$

have, for each (x_0, ν_0, u, v) in a neighborhood of $(\bar{x}_0, \bar{\nu}_0, 0_u, \bar{v})$, a unique solution, $\mu_0=M_1(x_0, \nu_0, u, v)$, of class C'.

This result follows at once from the implicit function theorems

* See paper IV for the definitions of terms such as composite, neighborhood, region, differential.

mentioned above,* because of assumptions C and the fact that $(\mu_0, x_0, \nu_0, u, v) = (\bar{\mu}_0, \bar{x}_0, \bar{\nu}_0, 0_u, \bar{v})$ is an initial solution of these equations.

COROLLARY 2.2. *For each (x, ν, U, V) the equations*

$$\Phi_1(x, \mu, \nu) = U,$$

$$\Phi_2(x, \mu, \nu) = V$$

have a unique continuous solution $\mu = M_2(x, \nu, U, V)$ which is equal to $dM_1(\bar{x}_0, \bar{\nu}_0, 0_u, \bar{v}; x, \nu, U, V)$.

The existence and unicity of the solution is an immediate consequence of assumption C; and the last assertion in the corollary can be established by setting $x_0 = \bar{x}_0 + ax$, $\nu_0 = \bar{\nu}_0 + a\nu$, $u = aU$, $v = \bar{v} + aV$, and $\mu_0 = M_1(x_0, \nu_0, u, v)$ in equations (2.1) and differentiating the resulting identities in a at $a = 0$.

Then to obtain the desired imbedding theorems we make a final assumption.

E. Either the function M_1, mentioned above, is independent of ν_0, or the transformation

$$\mu - d_\nu M_1[\bar{x}_0, \bar{\nu}_0, 0_u, \bar{v}; d_\mu n(\bar{x}_0, \bar{\mu}_0; \mu)]$$

has a linear continuous reciprocal.

It follows readily from Lemma 2.1 that if M_1 is independent of ν_0 and if $M(x_0, u, v)$ is defined to be $M_1(x_0, \nu_0, u, v)$, then the functions $\phi_1[x_0, M(x_0, u, v), \nu_0]$, $\phi_2[x_0, M(x_0, u, v), \nu_0]$ are independent of ν_0, at least for (x_0, ν_0, u, v) sufficiently near to $(\bar{x}_0, \bar{\nu}_0, 0_u, \bar{v})$.

THEOREM 1. *In a sufficiently small neighborhood of $\bar{\mu}_0$ the equations*

$$\phi_1[x_0, \mu_0, n(x_0, \mu_0)] = u,$$

$$\phi_2[x_0, \mu_0, n(x_0, \mu_0)] = v$$

possess a unique solution $\mu_0 = M(x_0, u, v)$ of class C' for (x_0, u, v) in some neighborhood of $(\bar{x}_0, 0_u, \bar{v})$.

From what has been said above it suffices to consider the case where M_1 is not independent of ν_0. Hence the equation

$$\mu_0 = M_1[x_0, n(x_0, \mu_0), u, v]$$

has a solution $\mu_0 = M(x_0, u, v)$ which is effective in the theorem.

Similarly a consideration of the equation

$$\mu = dM_1[\bar{x}_0, \bar{\nu}_0, 0_u, \bar{v}; x, N(x, \mu), U, V]$$

* See paper I, p. 65, and paper IV, p. 150.

yields the following corollary:

COROLLARY 2.3. *For each* (x, U, V) *the equations*

$$\Phi_1[x, \mu, N(x, \mu)] = U,$$
$$\Phi_2[x, \mu, N(x, \mu)] = V$$

have a unique continuous solution $\mu = M_3(x, U, V) = dM(\bar{x}_0, 0_u, \bar{v}; x, U, V)$.

COROLLARY 2.4. *Let* s *be an arbitrary integer,* $(x_1, \cdots, x_s)$ *a set of elements of* $\mathfrak{X}$, *and* $(\mu_1, \cdots, \mu_s)$ *a set of elements of* $\mathfrak{M}$ *satisfying* $\Phi_1[x, \mu, N(x, \mu)] = 0_u$. *Then there are functions* $x_0(b)$ *and* $\mu_0(b)$ *of class* C' *for values of* $b = (b_1, \cdots, b_s)$ *sufficiently near to* $(0, \cdots, 0)$, *which satisfy the equation*

$$\phi_1[x_0, \mu_0, n(x_0, \mu_0)] = 0_u,$$

$x_0(0) = x_0$, $\mu_0(0) = \bar{\mu}_0$, *and* $\partial x_0(0)/\partial b_i = x_i$, $\partial \mu_0(0)/\partial b_i = \mu_i$ *for* $1 \leqq i \leqq s$.

Define V_i to be $\Phi_2[x_i, \mu_i, N(x_i, \mu_i)]$, $(i = 1, \cdots, s)$, $x_0(b) = x_0(b_1, \cdots, b_s) = \bar{x}_0 + b_i x_i$, and $\mu_0(b)$ to be $M[x_0(b), 0_u, \bar{v} + b_i V_i]$, where a repeated index indicates a sum. Then the first part of the corollary is obvious, and the latter part is an easy consequence of corollary 2.3.

3. **The multiplier rule.** The usual mode of proof utilized in the calculus of variations suffices to establish the following lemma:

LEMMA 3.1. *There are constants* $l_0, c_1, \cdots, c_q$, *not all zero, such that*

$$g(x, \mu) = l_0 df(\bar{x}_0, \bar{\mu}_0; x, \mu) + c_k d\psi_k(\bar{x}_0, \bar{\mu}_0; x, \mu) = 0$$

for every (x, μ) *satisfying the equation*

$$\Phi_1[x, \mu, N(x, \mu)] = 0_u.$$

To simplify the succeeding demonstration we shall define $g_1(x)$ and $g_2(\mu)$ to be $l_0 d_x f(\bar{x}_0, \bar{\mu}_0; x) + c_k d_x \psi_k(\bar{x}_0, \bar{\mu}_0; x)$ and $l_0 d_\mu f(\bar{x}_0, \bar{\mu}_0; \mu) + c_k d_\mu \psi_k(\bar{x}_0, \bar{\mu}_0; \mu)$, respectively. It is then obvious that $g(x, \mu) = g_1(x) + g_2(\mu)$ and that g_1 and g_2 are linear continuous functionals on $\mathfrak{X}$ and $\mathfrak{M}$, respectively.

LEMMA 3.2. *The functions* $\Phi_1[x, M_3(x, U, V)]$, $N[x, M_3(x, U, V)]$, *and* $g[x, M_3(x, U, V)]$ *are independent of* x *and* V; *the latter is a linear continuous functional* J *on* $\mathfrak{U}$; *and for every* x

$$g_1(x) + g_2[M_3(x, 0_u, 0_v)] = 0. \tag{3.1}$$

It follows at once from Corollary 2.3 that Φ_1 is independent of x and V. Hence we have $\Phi_1[\bar{x}, \bar{\mu}, N(\bar{x}, \bar{\mu})] = 0_u$, where $\bar{x} = x_1 - x_2$, and $\bar{\mu} = M_3(\bar{x}, 0_u, V_1 - V_2)$. Consequently we have, by Lemma 3.1, $g(\bar{x}, \bar{\mu}) = 0$ for every x_1, x_2, V_1, and V_2; this proves the rest of the lemma, since $M_3(x_1, U, V_1) - M_3(x_2, U, V_2) = M_3(x_1 - x_2, 0_u, V_1 - V_2)$.

To complete the proof of the multiplier rule we shall now prove that there is a linear continuous operation $u = P(\mu)$ on $\mathfrak{M}$ to $\mathfrak{U}$ such that $g_2(\mu) = J[P(\mu)]$ for every μ. To this end define $P(\mu)$ to be $\Phi_1[0_x, \mu, N(0_x, \mu)]$ and $Q(\mu)$ to be $\Phi_2[0_x, \mu, N(0_x, \mu)]$. Therefore, by Corollary 2.3, $\mu = M_3[0_x, P(\mu), Q(\mu)]$; as a consequence it is evident that $g_2(\mu) = g_2(M_3[0_x, P(\mu), Q(\mu)])$ which, by Lemma 3.2, is $J[P(\mu)]$.

THEOREM 2. *There exist constants $l_0, c_1, \cdots, c_q$, not all zero, and a linear, continuous, and real-valued functional J on $\mathfrak{U}$ such that*

$$(1) \qquad l_0 d_\mu f(\bar{x}_0, \bar{\mu}_0; \mu) + c_k d_\mu \psi_k(\bar{x}_0, \bar{\mu}_0; \mu) - J(\Phi_1[0_x, \mu, N(0_x, \mu)]) = 0$$

for every μ in $\mathfrak{M}$, and

$$(2) \qquad l_0 d_x f(\bar{x}_0, \bar{\mu}_0; x) + c_k d_x \psi_k(\bar{x}_0, \bar{\mu}_0; x) - J(\Phi_1[x, 0_\mu, N(x, 0_\mu)]) = 0$$

for every x in $\mathfrak{X}$.

From the previous discussion it is evident that

$$g_2(\mu) = l_0 d_\mu f(\bar{x}_0, \bar{\mu}_0; \mu) + c_k d_\mu \psi_k(\bar{x}_0, \bar{\mu}_0; \mu) = J[P(\mu)]$$

and that $P(\mu)$ is $\Phi_1[0_x, \mu, N(0_x, \mu)]$. Equation (2), the "transversality condition," follows from equation (3.1). For,

$$g_2[M_3(x, 0_u, 0_v)] = J(P[M_3(x, 0_u, 0_v)]).$$

Then making use, on the one hand, of the definitions of P and Φ_1, and on the other hand of Corollary 2.3, we obtain the desired conclusion.

4. **Instances of the theory.** We shall illustrate the generality of our theory by working out two special cases, the minima of functions of several variables and the problem of Bolza in the calculus of variations. However the theory also includes the situations treated by Hahn, Graves, and the author, as well as many others.

Consider first the problem of minimizing a real-valued function $g(x_1, \cdots, x_m)$ of m real variables in the class of points $x_0 = (x_1, \cdots, x_m)$ satisfying equations $h_k(x_0) = 0$, $(k = 1, \cdots, q)$. It is assumed that g has a minimum at a point $\bar{x}_0$ in this class and that g and h_k have first differentials at this point. Then to indicate the applicability of our theory to this case we choose the sets $\mathfrak{M}$, $\mathfrak{N}$, $\mathfrak{M}_0$, $\mathfrak{N}_0$, $\mathfrak{U}$, and $\mathfrak{B}$ to be equal to the set of real numbers, $\mathfrak{X}$ to be the m-dimensional euclidean space, and $\mathfrak{X}_0$ to be the region of definition of

the functions g and $h_1, \cdots, h_q$. We then make the following definitions: $\psi_k(x_0, \mu_0) \equiv h_k(x_0)$, $(k=1, \cdots, q)$, $f(x_0, \mu_0) \equiv g(x_0)$, $n(x_0, \mu_0) \equiv 0$, $\phi_1(x_0, \mu_0, \nu_0) \equiv \mu_0$, and $\bar{\mu}_0 \equiv 0$. Then it is clear that $f(x_0, \mu_0)$ is a minimum at $(\bar{x}_0, \bar{\mu}_0)$ in the class of all points (x_0, μ_0) in the region which satisfy

$$\psi_k(x_0, \mu_0) = 0, \qquad k = 1, \cdots, q,$$
$$\mu_0 = 0,$$

and all previous assumptions are satisfied.

Our multiplier rule then implies the existence of constants $l_0, c_1, \cdots, c_q$, not all zero, and a constant e such that

$$l_0 g_{x_i}(\bar{x}_0)dx_i + c_k h_{kx_i}(\bar{x}_0)dx_i = 0, \qquad ed\mu = 0$$

for every $dx_1, \cdots, dx_m$ and every $d\mu$; therefore we have $e=0$ and

$$l_0 g_{x_i}(\bar{x}_0) + c_k h_{x_i}(\bar{x}_0) = 0, \qquad i = 1, \cdots, m,$$

which is the familiar form of the multiplier rule for this problem. The problem of minimizing a function f on a region $\mathfrak{X}_0$ of a normed linear space in the class of points satisfying functional equations $\psi_k(x_0)=0$, $(k=1, \cdots, q)$, can be treated in precisely the manner indicated above.

Let us turn our attention to the problem of Bolza in the calculus of variations, where, for simplicity, we shall suppose that the admissible arcs are of class C'. Here we are given an arc $\bar{y}_1(x), \cdots, \bar{y}_s(x)$, $(\bar{x}_1 \leqq x \leqq \bar{x}_2)$, which minimizes a functional

$$I[x_1, x_2, y, y'] = g[x_1, y(x_1), x_2, y(x_2)] + \int_{x_1}^{x_2} h[x, y(x), y'(x)]dx$$

in the class of admissible arcs which satisfy certain side conditions

$$\phi_\alpha[x, y(x), y'(x)] = 0, \qquad \alpha = 1, \cdots, r < s,$$

and end conditions

$$\psi_k[x_1, y(x_1), x_2, y(x_2)] = 0, \qquad k = 1, \cdots, q.$$

For our purposes it is sufficient to restrict attention to the class $\mathfrak{N}_0$ of all admissible sets of functions $(y_1(x), \cdots, y_s(x))$, $(\bar{x}_1 - e \leqq x \leqq \bar{x}_2 + e)$, where e is some fixed positive constant. The class which is composed of all sets $(y_1'(x), \cdots, y_s'(x))$ of derivatives of an element $\nu_0 = (y_1(x), \cdots, y_s(x))$ is $\mathfrak{M}_0$. Then if the transformation

$$y_i(x) = y_i(\bar{x}_1) + \int_{\bar{x}_1}^{x} y_i'(t)dt \tag{4.1}$$

is performed upon the functionals I and ψ_k, they become functionals $f(x_0, \mu_0)$ and $\psi_k(x_0, \mu_0)$ of $x_0=(x_1, x_2, y(\bar{x}_1))$, $\mu_0=(y')$. Thus $\mathfrak{X}_0$ is the set of all $(x_1, x_2, y_{01}, \cdots, y_{0s})$, where $|\bar{x}_1-x_1|<e$, $|\bar{x}_2-x_2|<e$, and $(y_{01}, \cdots, y_{0s})$ is the value of an admissible set at $x=\bar{x}_1$.

To complete our transformation of the problem of Bolza we define $\phi_1(x_0, \mu_0, \nu_0)$ to be the set of m functions

$$\phi_1[x, \nu_0(x), \mu_0(x)], \cdots, \phi_m[x, \nu_0(x), \mu_0(x)]$$

and $n(x_0, \mu_0)$ to be the set of s functions of (x_0, μ_0) which appear in the right members of equations (4.1). It is then a simple matter to verify assumptions A through D of §1, if it is assumed that the matrix $\phi_{\alpha y_i'}$ of partial derivatives has rank m along the minimizing arc. Assumption E can be verified without difficulty by making use of the theory of Volterra integral equations.

The functional J appearing in our multiplier rule can be shown* to be representable as

$$-\int_{\bar{x}_1}^{\bar{x}_2} u_\alpha(x)\lambda_\alpha(x)dx.$$

Now, if we define F to be $l_0h+\lambda_\alpha\phi_\alpha$, identity (1) of Theorem 2 yields the following equations:

$$F_{y_i}(x)dx + l_0(g_{y_{i1}} + g_{y_{i2}}) + c_k(\psi_{ky_{i1}} + \psi_{ky_{i2}}) = 0,$$

$$(-1)^bF(\bar{x}_b) + l_0[g_{x_b} + g_{y_{ib}}y_i'(\bar{x}_b)] + c_k[\psi_{kx_b} + \psi_{ky_{ib}}y_i'(\bar{x}_b)] = 0, \qquad b = 1, 2;$$

moreover identity (2) becomes

$$F_{y_i'}(x) - \int_{\bar{x}_2}^{x} F_{y_i}(t)dt + l_0g_{y_{i2}} + c_k\psi_{ky_{i2}} = 0.$$

By very obvious manipulations these equations can be put into the more familiar form

$$[(F - y_i'F_{y_i'})dx + F_{y_i'}dy_i]_1^2 + l_0dg + c_kd\psi_k = 0,$$

$$F_{y_i'} = \int_{\bar{x}_1}^{x} F_{y_i}dx + d_i,$$

which is the multiplier rule for the problem of Bolza.

THE UNIVERSITY OF CHICAGO

* See Hahn, loc. cit., pp. 544–546.

ON THE nTH DERIVATIVE OF $f(x)$*

H. S. WALL

Let $y_1, y_2, y_3, \cdots$ be defined recursively as follows: y_1 is the logarithmic derivative of a function $y=f(x)$, and $y_\nu = D_x y_{\nu-1}$, $(\nu=2, 3, 4, \cdots)$. Then the successive derivatives $y', y'', y''', \cdots$ of y with respect to x are polynomials in y and the y_ν. In fact, $y'=yy_1$, $y''=y(y_2+y_1^2)$, $y'''=y(y_3+3y_1y_2+y_1^3)$, and

$$y^{(n)} = y \sum A^{(n)}_{\nu_1\nu_2\cdots\nu_n} y_1^{\nu_1} y_2^{\nu_2} \cdots y_n^{\nu_n}, \tag{1}$$

where $A^{(n)}_{\nu_1\nu_2\cdots\nu_n}$ is a positive integer and the summation is taken for all non-negative integral solutions $\nu_1, \nu_2, \nu_3, \cdots, \nu_n$ of the equation

$$\nu_1 + 2\nu_2 + 3\nu_3 + \cdots + n\nu_n = n. \tag{2}$$

This statement may readily be proved by mathematical induction. The principal object of the present note is to prove the following theorem:

THEOREM. *The integer $A^{(n)}_{\nu_1\nu_2\cdots\nu_n}$ in* (1) *is equal to the number of ways that n different objects can be placed in compartments, one in each of ν_1 compartments, two in each of ν_2 compartments, three in each of ν_3 compartments, $\cdots$, without regard to the order of arrangement of the compartments.*

1. Generalized binomial coefficients. Let $k, m, n, (kn \leqq m)$, be positive integers, and denote by $C^{(k)}_{m,n}$ the number of ways that kn objects can be selected from m objects and placed in n compartments, k in each compartment, where no account is taken of the order of arrangement of the compartments. Thus $C^{(1)}_{m,n}$ is the binomial coefficient $C_{m,n}=m!/[n!(m-n)!]$. We have

$$n!\cdot C^{(k)}_{m,n} = C_{m,kn}\cdot(C_{kn,k}\cdot C_{k(n-1),k}\cdot \cdots \cdot C_{k,k}),$$

or

$$C^{(k)}_{m,n} = m!/[n!(m-kn)!(k!)^n]. \tag{3}$$

This has meaning if $m \geqq kn$. For special 0 values of the indices we shall consider $C^{(k)}_{m,n}$ to be defined by (3) by taking $0!=1$. Thus if $k \geqq 0$, $m \geqq 0$, we have $C^{(k)}_{m,0}=1$.

* Presented to the Society, September 5, 1936.

If (2) holds, it will be seen that

$$(4)\qquad C^{(1)}_{k_1,\nu_1}\cdot C^{(2)}_{k_2,\nu_2}\cdot\cdots\cdot C^{(n)}_{k_n,\nu_n}=\frac{n!}{\nu_1!\nu_2!\cdots\nu_n!(1!)^{\nu_1}(2!)^{\nu_2}\cdots(n!)^{\nu_n}},$$

where $k_1=n$, $k_2=n-\nu_1$, $k_3=n-\nu_1-2\nu_2$, $\cdots$. We are to prove that this is the value of $A^{(n)}_{\nu_1\nu_2\cdots\nu_n}$ in (1). For the proof we need the following identities which will be seen to hold for all values of m, n, k for which the symbols involved have been defined:

$$(5)\qquad C^{(k+1)}_{m,n}=C^{(k+1)}_{m-1,n}+C^{(k)}_{m-1,1}\cdot C^{(k+1)}_{m-k-1,n-1},$$

$$(6)\qquad (n+1)\cdot C^{(k)}_{m,n+1}=C^{(k)}_{m,n}\cdot C^{(k)}_{m-kn,1}.$$

Let it be remarked in passing that if $P_{n,k}=C^{(k)}_{n,0}+C^{(k)}_{n,1}x+C^{(k)}_{n,2}x^2+\cdots$, then from (5) it follows that $P_{n,k}=P_{n-1,k}+C_{n-1,k-1}xP_{n-k,k}$. Also $P'_{n,k}=C_{n,k}P_{n-k,k}$.

2. **Derivation of the formula for $A^{(n)}_{\nu_1\nu_2}\cdots_{\nu_n}$.** Denote the sum in (1) by S_n, and write S_n as a polynomial in y_1, $S_n=\sum_{\nu=0}^{n}S^{(1)}_{n,\nu}y_1^{\nu}$, where $S^{(1)}_{n,\nu}$ is independent of y_1. We begin by showing that

$$(7)\qquad S^{(1)}_{i,j}=C^{(1)}_{i,j}S^{(1)}_{i-j,0},\qquad 0\leqq j\leqq i.$$

We use induction on the subscript difference $i-j=k$. From the relation $yS_n=D_x[yS_{n-1}]$ it follows that

$$(8)\qquad S^{(1)}_{n,\nu}=S^{(1)}_{n-1,\nu-1}+(1+\nu)y_2S^{(1)}_{n-1,\nu+1}+D_xS^{(1)}_{n-1,\nu},\qquad \nu=0,1,2,\cdots,n,$$

with the agreement that $S_{i,j}=0$ if $j<0$ or $j>i$. Assuming that (7) holds for $k<q$ we shall prove that it holds for $k=q$. Accordingly, we choose n, ν, $(0\leqq\nu<n)$, in (8) so that $n-\nu=q$. Then, by our assumption, (8) may be written in the form

$$(9)\qquad S^{(1)}_{n,\nu}=\begin{cases}S^{(1)}_{n-1,\nu-1}+(1+\nu)y_2C^{(1)}_{n-1,\nu+1}S^{(1)}_{n-\nu-2,0}+C^{(1)}_{n-1,\nu}D_xS^{(1)}_{n-\nu-1,0}, & \text{if } q>1;\\ S^{(1)}_{n-1,\nu-1}+C^{(1)}_{n-1,\nu}D_xS^{(1)}_{n-\nu-1,0}, & \text{if } q=1.\end{cases}$$

Replace n by $n-\nu$, and ν by 0 in (8), and eliminate $D_xS^{(1)}_{n-\nu-1,0}$ in (9). The result, by (6), is

$$(10)\qquad S^{(1)}_{n,\nu}=S^{(1)}_{n-1,\nu-1}+C^{(1)}_{n-1,\nu}S^{(1)}_{n-\nu,0}.$$

Hence

$$S_{n,\nu}^{(1)} = \sum_{i=0}^{\nu} [S_{n-i,\nu-i}^{(1)} - S_{n-i-1,\nu-i-1}^{(1)}] = \left[\sum_{i=0}^{\nu} C_{n-i-1,\nu-i}^{(1)} \right] S_{n-\nu,0}^{(1)};$$

or, by (5) with $k=0$, $S_{n,\nu}^{(1)} = C_{n,\nu}^{(1)} S_{n-\nu,0}$, as was to be proved.

We next put

$$S_{m,0}^{(p-1)} = \sum_{\nu=0}^{[m/p]} S_{m,\nu}^{(p)} y_p^{\nu}, \qquad p = 2, 3, 4, \cdots .$$

Then the formulas

$$S_{m,\nu}^{(p)} = C_{m,\nu}^{(p)} S_{m-p\nu,0}^{(p)}, \tag{11}$$

$$S_{m,\nu}^{(p)} = C_{m-1,1}^{(p-1)} S_{m-p,\nu-1}^{(p)} + (1+\nu) y_{p+1} S_{m-1,\nu+1}^{(p)} + D_z S_{m-1,\nu}^{(p)} \tag{12}$$

hold for $p=1$. Assuming that they hold for $p<k$, $k>1$, we may then prove them for $p=k$. To do this, put $\nu=0$ and $p=k-1$ in (12), and equate coefficients of like powers of y_k. The result is the equation (12) with $p=k$. Thus (12) is true when $p=k$, and in particular $S_{k\nu,\nu}^{(k)} = C_{k\nu-1,1}^{(k-1)} S_{k(\nu-1),\nu-1}^{(k)}$. Hence by (5) we find that $S_{k\nu,\nu}^{(k)} = C_{k\nu,\nu}^{(k)} S_{0,0}^{(k)}$; so that (11) holds for $p=k$ provided $m-k\nu=0$. The proof of (11) for $p=k$ may now be carried out along the lines of the proof of (7). with induction, in this case, on the difference $m-k\nu$.

After (11) has been proved it follows at once by (4) that

$$A_{\nu_1\nu_2\cdots\nu_n}^{(n)} = \frac{n!}{\nu_1!\nu_2!\cdots\nu_n!(1!)^{\nu_1}(2!)^{\nu_2}\cdots(n!)^{\nu_n}}. \tag{13}$$

3. **Application.** In conclusion I shall give examples to illustrate the application of the foregoing result.

EXAMPLE 1. Let $a_{n,k}$ denote the number of ways that n different objects can be distributed among k compartments, where no account is taken of the order of arrangement of the compartments, and at least one object is placed in each compartment. Then elementary considerations will show that $a_{n,k} = ka_{n-1,k} + a_{n-1,k-1}$. Put

$$y = e^{te^z} = e^t \sum_{\nu=0}^{\infty} L_\nu(t) x^\nu/\nu!.$$

Then by (1) we find that $L_n(t) = a_{n,1}t + a_{n,2}t^2 + \cdots + a_{n,n}t^n$.

Put $g_k(y) = \sum_{\nu=0}^{\infty} a_{k+\nu,k} y^\nu$. It follows that $g_1(y) = 1/(1-y)$, $g_k(y) = g_{k-1}(y)/(1-ky)$, $(k=2, 3, 4, \cdots)$, or $g_k(y) = 1/[(1-y)(1-2y)\cdots(1-ky)]$. Hence $L_n(1)$, the number of ways that n different ob-

jects can be distributed among n or fewer compartments, is the coefficient of y^n in the power series $P(y)$ for the function

$$\frac{y}{(1-y)}+\frac{y^2}{(1-y)(1-2y)}+\cdots+\frac{y^m}{(1-y)(1-2y)\cdots(1-my)},$$

where $m \geqq n$. The number $L_n(1)$ is also the coefficient of $x^n/n!$ in the power series for the function $e^{(e^x-1)}$.

The $a_{n,k}$ are given explicitly by the formula

$$a_{n,k}=\frac{(-1)^{k+1}}{k!}\sum_{\nu=0}^{k} C_{k,\nu}(-1)^{\nu+1}\nu^n.$$

EXAMPLE 2. Put $y=(1+x)^{-t}$ in (1) and then set $x=0$. There results this identity:

$$\frac{t(t+1)(t+2)\cdots(t+n-1)}{n!} = \sum \frac{t^{\nu_1+\nu_2+\cdots+\nu_n}}{(1\cdot 2\cdot\cdots\cdot\nu_1)(2\cdot 4\cdot\cdots\cdot 2\nu_2)\cdots(n\cdot 2n\cdot\cdots\cdot n\nu_n)}, \tag{14}$$

where the summation is taken as in (1). On putting $t=1$ in (14) we obtain the following theorem:*

THEOREM. *Form a partition of n by taking at most one integer from each of the progressions* $1, 2, 3, \cdots$; $2, 4, 6, \cdots$; $3, 6, 9, \cdots$; $\cdots$. *Multiply together the terms of each progression up to and including the integer chosen. Let the products so formed be $a, b, c, \cdots$. Then* $\sum[1/(a\cdot b\cdot c\cdot\cdots)]=1$, *where the sum is taken for all such partitions of n.*

EXAMPLE 3. If we differentiate the members of (14) with respect to t and then set $t=1$, we get the formula

$$1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n} = \sum \frac{(\nu_1+\nu_2+\cdots+\nu_n)}{(1\cdot 2\cdot\cdots\cdot\nu_1)(2\cdot 4\cdot\cdots\cdot 2\nu_2)\cdots(n\cdot 2n\cdot\cdots\cdot n\nu_n)}.$$

This may likewise be interpreted as a theorem on partitions of n.

NORTHWESTERN UNIVERSITY

* Jacobi, *Zur combinatorischen Analysis*, Crelle's Journal, vol. 22 (1841), pp. 372–374.

THE GEOMETRY OF THE WHIRL-MOTION GROUP G_6: ELEMENTARY INVARIANTS*

EDWARD KASNER AND JOHN DE CICCO

In this paper we shall study the elementary geometry of the oriented lineal elements of the plane with respect to the whirl-motion group G_6. We give results in addition to those found in a paper by Kasner, *The group of turns and slides and the geometry of turbines*, published in 1911 in the American Journal of Mathematics (vol. 33, pp. 193–202) and the paper by the authors, *The geometry of turbines, flat fields, and differential equations*, published in 1937 in the American Journal of Mathematics (vol. 59, pp. 545–563). The present paper can be read independently of the two earlier papers and contains the foundation of a new geometry of differential elements.

We begin by considering certain simple operations or transformations on the oriented lineal elements of the plane. A turn T_α converts each element into one having the same point and a direction making a fixed angle α with the original direction. By a slide S_k the line of the element remains the same and the point moves along the line a fixed distance k. These transformations together generate a continuous group of three parameters which we call the group of *whirl* transformations. It is noted that the group of whirls is isomorphic to the group of motions. These two three-parameter groups are commutative and together generate a new group of six parameters which we term the *whirl-motion group* G_6. It is our purpose to study the simple geometry of this group G_6. We find the fundamental invariants.

We define ∞^1 elements to be a *series* of elements; this includes a union or curve as a special case. A set of ∞^2 elements is said to form a *field* of elements, which corresponds to a differential equation of the first order $F(x, y, y') = 0$. A *turbine* is the series which is obtained by applying a turn T_α to the elements of an oriented circle (which may be a null circle). A *flat field* is the field that is obtained from the totality of all elements which are determined by the set of all oriented circles containing a given element. We obtain the elementary relationships between elements, turbines, and flat fields.

For the analytic representation, it will be convenient to define an element by the coordinates (u, v, w) where v is the perpendicular from the origin, u is the angle between the perpendicular and the initial

* Presented to the Society, December 30, 1937.

line, and w is the distance between the foot of the perpendicular and the point of the element.

1. **The whirl-motion group G_6 of transformations.*** The equations of the turn T_α are

$$(1) \quad \bar{u} = u + \alpha, \ \bar{v} = v \cos \alpha + w \sin \alpha, \ \bar{w} = -v \sin \alpha + w \cos \alpha.$$

The equations of the slide S_k are

$$(2) \qquad \bar{u} = u, \qquad \bar{v} = v, \qquad \bar{w} = w + k.$$

Since any whirl may be expressed in the form $T_\beta S_k T_\alpha$, the equations of any whirl are

$$(3) \quad \begin{aligned} &\bar{u} = u + \alpha + \beta, \quad \bar{v} = v \cos (\alpha + \beta) + w \sin (\alpha + \beta) + k \sin \beta, \\ &\bar{w} = -v \sin (\alpha + \beta) + w \cos (\alpha + \beta) + k \cos \beta. \end{aligned}$$

The only contact transformations of the group of whirl transformations are

$$T_{-\pi/2} S_d T_{\pi/2}, \qquad T_{\pi/2} S_d T_{\pi/2}.$$

The first represents a dilatation D_d and the second, which may be written $T_\pi D_d$, represents a dilatation accompanied by a reversal of orientation.

It is then seen that any whirl transformation may be given by

$$S_k D_d T_\alpha;$$

so that the equations of any whirl transformation may be put in the form

$$(4) \quad \begin{aligned} &\bar{u} = u + \alpha, \qquad \bar{v} = v \cos \alpha + w \sin \alpha + d, \\ &\bar{w} = -v \sin \alpha + w \cos \alpha + k. \end{aligned}$$

It is observed that the group of whirl transformations and the group of motions are isomorphic. These two groups are commutative and together generate a continuous group of six parameters which we call the whirl-motion group G_6.

The equations of any motion are

$$(5) \quad \begin{aligned} &\bar{u} = u + \lambda, \qquad \bar{v} = v + \mu \cos u + \nu \sin u, \\ &\bar{w} = w - \mu \sin u + \nu \cos u. \end{aligned}$$

From (4) and (5), we find that the equations of any whirl-motion transformation are

* See Kasner, *The group of turns and slides and the geometry of turbines*, loc. cit.

$$
\begin{aligned}
\bar{u} &= u + \alpha + \lambda, \\
\bar{v} &= v \cos \alpha + w \sin \alpha + \mu \cos (u + \alpha) + \nu \sin (u + \alpha) + d, \\
\bar{w} &= -v \sin \alpha + w \cos \alpha - \mu \sin (u + \alpha) + \nu \cos (u + \alpha) + k.
\end{aligned} \tag{6}
$$

This is our fundamental six-parameter group G_6, containing the whirl group W_3, and the motion group M_3 as three-parameter subgroups.

2. **The invariants for two elements.** The *angle* between two elements is the angle between their lines. The *distance* between two parallel elements is the distance between their points.

THEOREM 1. *Under G_6 we find that* (1) *two non-parallel elements possess the angle between them as the unique invariant, and that* (2) *parallel elements are transformed into parallel elements in such a way that the distance between them is preserved.*

3. **The turbine.** A *turbine* is the series which is obtained by applying a turn T_a to the elements of an oriented circle.

The equations of a turbine are

$$v = a \cos u + b \sin u + r, \qquad w = -a \sin u + b \cos u + s.$$

We call (a, b, r, s) a set of turbine coordinates.

Under G_6, any turbine is converted into a turbine. Thus G_6 induces a correspondence between the turbines of the plane. By (6) we find that this correspondence is given by the equations

$$
\begin{aligned}
\bar{a} &= a \cos \lambda - b \sin \lambda + \mu \cos \lambda - \nu \sin \lambda, \\
\bar{b} &= a \sin \lambda + b \cos \lambda + \mu \sin \lambda + \nu \cos \lambda, \\
\bar{r} &= r \cos \alpha + s \sin \alpha + d, \\
\bar{s} &= -r \sin \alpha + s \cos \alpha + k.
\end{aligned} \tag{7}
$$

From the equations of a turbine, we obtain the following theorem:

THEOREM 2. *Two non-parallel elements determine a unique turbine.*

THEOREM 3. *The necessary and sufficient condition that two distinct turbines possess a common element (in which case they have only one common element) is*

$$(a_2 - a_1)^2 + (b_2 - b_1)^2 = (r_2 - r_1)^2 + (s_2 - s_1)^2.$$

THEOREM 4. *Two turbines have two and only two invariants, namely,*

$$(a_2 - a_1)^2 + (b_2 - b_1)^2 \quad \text{and} \quad (r_2 - r_1)^2 + (s_2 - s_1)^2.$$

From this and Theorem 3, it follows that two intersecting turbines

possess the squared distance between their centers as the unique invariant.

THEOREM 5. *A given element and a given turbine possess the unique invariant*

$$[a-(v-r)\cos u+(w-s)\sin u]^2+[b-(v-r)\sin u-(w-s)\cos u]^2.$$

From this it is seen that a given element E and a given turbine T possess as the unique invariant the squared distance between the element E and that element on T which is parallel to E.

4. **Conjugate turbines.** Two turbines $\overline{T}$ and T are said to be *conjugate* if they have the same circle as point locus and the elements of the two turbines are symmetrically related to the elements of the circle.

The two turbines $\overline{T}(\bar{a}, \bar{b}, \bar{r}, \bar{s})$ and $T(a, b, r, s)$ are conjugate if and only if

$$\bar{a} = a, \qquad \bar{b} = b, \qquad \bar{r} = r, \qquad \bar{s} = -s.$$

THEOREM 6. *The conjugate turbines of two given turbines do or do not possess a common element according as the two given turbines do or do not possess a common element.*

5. **The flat field.** The totality of elements determined by the set of all circles which contain a given element is called a *flat field*. The given element is called the *center* (or *central element*) of the flat field.

The equation of a flat field is

$$w = -(v-\bar{v})\cot\tfrac{1}{2}(\bar{u}-u)-\bar{w},$$

where $(\bar{u}, \bar{v}, \bar{w})$ are the hessian coordinates of the central element.

Under the whirl-motion group G_6, any flat field is converted into a flat field. Thus G_6 induces a transformation between the flat fields of the plane. By equation (6), this induced correspondence is given by the equations

$$\begin{aligned}
\bar{\bar{u}} &= \bar{u}-\alpha+\lambda,\\
\bar{\bar{v}} &= \bar{v}\cos\alpha-\bar{w}\sin\alpha+\mu\cos(\bar{u}-\alpha)+\nu\sin(\bar{u}-\alpha)+d,\\
\bar{\bar{w}} &= \bar{v}\sin\alpha+\bar{w}\cos\alpha-\mu\sin(\bar{u}-\alpha)+\nu\cos(\bar{u}-\alpha)-k.
\end{aligned} \tag{8}$$

It is seen that (8) is a whirl-motion transformation. *We call it the conjugate whirl-motion transformation* $\overline{WM}$ *of the whirl-motion transformation* WM *given by equation* (6).

THEOREM 7. *Three elements which are not all on the same turbine and which do not all possess the same direction determine a unique flat field.*

This follows from the equation of a flat field given above.

THEOREM 8. *The turbines which are contained in a flat field are those whose conjugate turbines contain the central element of the flat field.*

THEOREM 9. *Two flat fields whose central elements are not parallel always have a common turbine. Three flat fields, no two of whose central elements are parallel, have a common element or else a common turbine.*

DEFINITIONS. The *angle* between two given flat fields is the angle between their central elements. Two flat fields are said to be *parallel* or *supplementary* (*anti-parallel*) according as the angle between them is 0 or π. The *distance* between two parallel flat fields is defined to be the distance between their central elements.

THEOREM 10. *Under G_6, we find that* (1) *two non-parallel flat fields possess the angle between them as the unique invariant, and that* (2) *two parallel flat fields are transformed into parallel flat fields in such a way that the distance between them is preserved.*

THEOREM 11. *A given element and a given flat field possess the unique invariant*

$$\left[(v - \overline{V}) \cos \frac{u - \overline{U}}{2} - (w + \overline{W}) \sin \frac{u - \overline{U}}{2}\right]^2.$$

From this, we find the following geometric interpretation of the invariant: Let E be the given element and F the given flat field. Let the central element of F be denoted by $\overline{G}$. On the oriented line of E, construct the element G which is in the flat field F. Let l be the line connecting the points of $\overline{G}$ and G. The square of the perpendicular distance from the point of E to the line l is our invariant.

THEOREM 12. *A given turbine and a given flat field possess the unique invariant*

$$[a-(\bar{v}-r) \cos \bar{u}+(\bar{w}+s) \sin \bar{u}]^2+[b-(\bar{v}-r) \sin \bar{u}-(\bar{w}+s) \cos \bar{u}]^2.$$

From this we find the following geometric interpretation of our invariant. Let T be the given turbine and F the given flat field, let $\overline{T}$ be the conjugate turbine of T, and let $\overline{G}$ be the central element of F. Then our invariant is the square of the distance from the element $\overline{G}$ to the turbine $\overline{T}$. (See construction in Theorem 5.)

COLUMBIA UNIVERSITY AND
BROOKLYN COLLEGE

A NOTE ON NUMBERS OF THE FORM $a^2+\alpha b^2+\beta c^2+\alpha\beta d^2$

ROBERT M. THRALL

It has long been known that the only sets of values of α, β which give rise to universal forms $a^2+\alpha b^2+\beta c^2+\alpha\beta d^2$ are $\alpha=1$, $\beta=1, 2, 3$; $\alpha=2$, $\beta=2, 3, 4, 5$. We give here a theorem from which the universal character of all these multiplicative universal forms can be readily established. The methods and arguments used are completely algebraic in character. This note is closely allied to R. D. Carmichael's paper, *Proof that every positive integer is a sum of four integral squares.** We use the formulas on the first page of his proof without recording them here.

We remark that every prime number p is a divisor of the form $a^2+\alpha b^2+\beta c^2+\alpha\beta d^2$ in which a, b, c, d are relatively prime. For if p is prime to α, a theorem from the theory of quadratic residues states that p is a divisor of $a^2+\alpha b^2+\beta$. If p is not prime to α, take $a=c=d=0$, $b=1$. In both cases the bases are seen to be relatively prime.

We now require that α, β be positive integers with $\beta\geqq\alpha>0$. The results obtainable also hold for $\alpha=0$, which case has been treated by R. D. Carmichael.†

THEOREM 1. *For $\alpha<3$ and every prime number $p>\beta$ there exists a positive integer $q\leqq[4\beta/(3-\alpha)]^{1/2}$ such that $p\ q=a^2+\alpha b^2+\beta c^2+\alpha\beta d^2$.*

For $p<5$ we verify the truth of the theorem directly. Henceforth consider $p\geqq 5$. Since $p>\beta\geqq\alpha$, from the above remark we have a q' such that $pq'=a^2+\alpha b^2+\beta$, where a and b may evidently be taken less than $p/2$. Hence, $pq'\leqq(1+\alpha)p^2/4+\beta$ or $q'\leqq(1+\alpha)p/4+\beta/p<3p/4+1$; that is, $q'<p$ if $p\geqq 5$, which is the case now under consideration.

Let q be the smallest positive integer such that $pq=a^2+\alpha b^2+\beta c^2+\alpha\beta d^2$. From the above we have $q<p$. Also a, b, c, d are relatively prime, for otherwise the square of their greatest common divisor would divide q, leaving pq_1 in the same form as pq with $q_1<q$ contrary to the hypothesis that q has the smallest value possible.

We now consider the forms in the parentheses of the formula that would correspond to Carmichael's (2):

---

* Duke Mathematical Journal, vol. 2 (1936), pp. 243–245.

† American Mathematical Monthly, vol. 44 (1937), pp. 81–86.

$$A_1 = ax - \alpha by - \beta cz - \alpha\beta dt,$$
$$A_2 = bx + ay + \beta dz - \beta ct,$$
$$A_3 = cx - \alpha dy + az + \alpha bt,$$
$$A_4 = dx + cy - bz + at.$$

The expression A_4 is invariant under any change of x, y, z, t of the form:

$$x' = x + \alpha_1 c + \alpha_2 b + \alpha_3 a,$$
$$y' = y - \alpha_1 d + \alpha_4 b + \alpha_5 a,$$
$$z' = z + \alpha_2 d + \alpha_4 c + \alpha_6 a,$$
$$t' = t - \alpha_3 d - \alpha_5 c + \alpha_6 b,$$

where the α's are arbitrary. Under this transformation A_3 becomes

$$A_3' = A_3 + \alpha_1(c^2 + \alpha d^2) + (\alpha_2 - \alpha\alpha_5)(cb + ad) + (\alpha_3 + \alpha_4)(ac - \alpha bd) + \alpha_6(a^2 + \alpha b^2).$$

We may choose $x=t=0$, $y=b$, $z=c$ so that $A_4=0$. If a prime p divides a number of the form $a^2+\alpha b^2$, (a, b relatively prime, $\alpha=1$ or 2), then p is itself of that form, hence $q=1\leqq[4\beta/(3-\alpha)]^{1/2}$. For other primes p the greatest common divisor δ of $a^2+\alpha b^2$ and $c^2+\alpha d^2$ must be 1. For if δ has a prime power factor p_1^r then $(a^2+\alpha b^2)/p_1^r = a_1^2+\alpha b_1^2$; $(c^2+\alpha d^2)/p_1^r = c_1^2+\alpha d_1^2$ (this follows from the theory of divisors of the forms a^2+b^2 and a^2+2b^2); and $p(q/p_1^r)$ has the form considered. We can therefore choose α_1, α_6 (taking $\alpha_2=\alpha_3=\alpha_4=\alpha_5=0$) so that $A_3'=1$. Then in (2)* take $\sigma=\rho=0$ and λ, μ such that the bases in the first two braces are numerically less than or equal to $q/2$. Now suppose that $q>[4\beta/(3-\alpha)]^{1/2}$; that is, that $\beta<q^2(3-\alpha)/4$. The form A is replaced by A' with the new choice of x', y', z', t', and

$$\beta \leqq A'q \leqq \frac{(\alpha+1)}{4}q^2 + \beta < \frac{(\alpha+1)}{4}q^2 + \frac{(3-\alpha)}{4}q^2 = q^2.$$

Hence $0<A'<q$; but in view of Carmichael's formula (6) this contradicts the minimal nature of q. Therefore, $q\leqq[4\beta/(3-\alpha)]^{1/2}$, and the theorem is proved.

We now give a corollary which is useful in some applications of the theorem.

COROLLARY 1. *If* $q\leqq 3$ *and* $\alpha+\beta+1<9$, *then* $q\leqq 2$.

For, referring to (2) with $q=3$, we see that by choosing x', y', z', t',

* References are to equations in Carmichael's proof.

$\mu, \lambda, \rho, \sigma$ as in the proof of Theorem 1 we have $3A' \leqq 1+\alpha+\beta<9$. Therefore $A'<3$. But A' is an integer, and if it is less than three this contradicts the minimal character of q.

A form $ax^2+by^2+cz^2+dt^2$ is called universal when it represents all positive integers. From the multiplicative property of the form $a^2+\alpha b^2+\beta c^2+\alpha\beta d^2$ we see that if it represents every prime number it is universal. If the number q, defined in the preceding theorem, is one, then the form does represent every prime and therefore every positive integer. We turn to the consideration of particular values for α, β. In the following paragraphs the theory of the forms $a^2+\beta c^2$ will be assumed.

The case $\alpha=\beta=1$ gives $q \leqq 2^{1/2}$. Hence $q=1$, and we have the following theorem:

THEOREM 2. *Every positive integer is a sum of four integral squares.*

That the forms $a^2+\alpha b^2+\beta c^2+\alpha\beta d^2$, with $(\alpha, \beta)=(1, 2), (1, 3), (2, 2), (2, 3), (2, 4)$ are universal follows readily by application of Theorem 1 and its corollary. We shall treat just one typical case, $\alpha=2, \beta=3$, among these.

If $\alpha=2, \beta=3$, we have $q \leqq (12)^{1/2}$ or $q \leqq 3$. But, since $\alpha+\beta+1=6<9$, the corollary applies and $q \leqq 2$. If $2p=a^2+2b^2+3c^2+6d^2$, then a and c are of like parity. If they are both even, $p=b^2+2(a/2)^2+3d^2+6(c/2)^2$. If they are both odd, by proper choice of sign for a and c we have $(a^2+3c^2)/4=\lambda^2+3\mu^2$, where $\lambda+3\mu=a$, $\lambda-\mu=c$, and $p=b^2+2\lambda^2+3d^2+6\mu^2$. We have now proved the following theorem:

THEOREM 3. *Every positive integer is representable in the form*

$$a^2+2b^2+3c^2+6d^2.$$

The case $\alpha=2, \beta=5$ gives $q \leqq (20)^{1/2}$ or $q \leqq 4$. But if $4p=a^2+2b^2+5c^2+10d^2$, (p an odd prime), then a and c must both be even, and $2p=b^2+2(a/2)^2+5d^2+10(c/2)^2$. Hence $q \leqq 3$. But, since $\alpha+\beta+1=8<9$, the corollary gives $q \leqq 2$, and

$$2p = a^2 + 2b^2 + 5c^2 + 10d^2, \tag{1}$$

where a and c must be of like parity. If they are even, $p=b^2+2(a/2)^2+5d^2+10(c/2)^2$. Hence suppose

$$a, c \text{ odd}. \tag{2}$$

Then by taking residues (mod 4) in (1),

$$b \equiv d \pmod 2. \tag{3}$$

Now by (1)

$$(4)^* \qquad p = \left(\frac{a+2b+5c+10d}{6}\right)^2 + 2\left(\frac{a-b+5c-5d}{6}\right)^2 + 5\left(\frac{a+2b-c-2d}{6}\right)^2 + 10\left(\frac{a-b-c+d}{6}\right)^2.$$

In view of (2) and (3), the numerators in (4) are all even. Then, unless exactly three of a, b, c, d are divisible by 3, we can choose signs for a, b, c, d so that

$$(5) \qquad a - b - c + d \equiv 0 \pmod 3.$$

Then all the other numerators in (4) are divisible by 3.

In the exceptional case either a and b or c and d are divisible by 3. But the identity

$$(6) \qquad 9(A^2 + 2B^2) = (A \pm 4B)^2 + 2(2A \mp B)^2$$

(repeated if necessary) shows that any multiple of 3 of the form x^2+2y^2 can be expressed in that form with x, y prime to 3. Then (5) can be verified as above, and $q = 1$. We have now proved the following theorem:

THEOREM 4. *Every positive integer is representable in the form*

$$a^2+2b^2+5c^2+10d^2.$$

UNIVERSITY OF ILLINOIS

A MOMENT-GENERATING FUNCTION WHICH IS USEFUL IN SOLVING CERTAIN MATCHING PROBLEMS†

EDWIN G. OLDS

1. **Introduction.** In a book published several years ago, Fry‡ devoted considerable attention to various aspects of a problem which he called, "the psychic research problem." His introductory problem is the following: "A spiritualistic medium claims to be able to tell the

---

* Formula (4) and the rest of the proof of this theorem were suggested by Gordon Pall.

† Presented to the Society and The Institute of Mathematical Statistics, December 30, 1937.

‡ T. C. Fry, *Probability and Its Engineering Uses*, Van Nostrand, New York, 1928, pp. 41–77.

color of a playing card without seeing it. In order to test her claims an experiment is conducted with four red and four black cards. These cards are thoroughly shuffled and placed face down on the table. The medium is told that there are four red and four black cards, but presumably knows nothing as to their arrangement. The experimenter picks up a card and, without either looking at it himself or showing it to the medium, asks its color. If she answers 'red,' he places it at one side of the table. If she answers 'black,' he places it on the other side of the table. This process is repeated until all cards are exhausted. If the medium does not have the ability which she claims to possess, what is the chance that there will be just one black card in the pile that should be red?"

Fry solves this problem by considering the number of chance orders of red and black which would match the order in which the medium calls the cards to the extent of producing exactly three red pairs. He also proposes and solves various other problems of this general type. A problem like Fry's, but concerned with tea-tasting, is thoroughly discussed by Fisher.*

Chapman† in 1931, contributed the solutions of several problems related to the expected number of correct matchings if a deck of t different cards is placed in correspondence with a given target deck of the same kind. One of his results is the function

$$p(s) = \frac{1}{s!}\left[\frac{1}{0!} - \frac{1}{1!} + \frac{1}{2!} - \cdots + (-1)^{t-s}\frac{1}{(t-s)!}\right],$$

which gives the probability of s correct matchings.

A more general problem has been considered by Greenwood.‡ He considers matching a target deck containing t sets of identical cards, with s_i cards in the ith set, $(i = 1, 2, \cdots, t)$. For the case where the number of cards is the same for each set, he obtains the mean and variance of the distribution of the numbers of correct matchings as s and $s^2(t-1)/(st-1)$, respectively. The purpose of the present paper is to develop a moment-generating function and derive the first four moments for the theoretical distribution of numbers of correct match-

* R. A. Fisher, *The Design of Experiments*, Oliver and Boyd, Edinburgh, 1935, chap. 2.

† D. W. Chapman, *The statistics of the method of correct matchings*, American Journal of Psychology, vol. 46 (1934), pp. 287–298.

‡ J. A. Greenwood, *Variance of a general matching problem*, Annals of Mathematical Statistics, vol. 9 (1938), pp. 56–59. See also M. S. Bartlett, *Properties of sufficiency and statistical tests*, Proceedings of the Royal Society, (A), vol. 160, pp. 268–282 (note especially the discussion of the $l \times m$ contingency table, pp. 278–280).

ings for this case where the number of cards is the same for each set.*

2. **Discussion of Fry's problem.** Before taking up the general case of t sets of s cards each, let us consider a solution of Fry's problem, which we quoted above.

Suppose the cards are tagged $R_1, R_2, R_3, R_4, B_1, B_2, B_3, B_4$. Let us consider the order in which the medium calls the cards as constituting the order in which a target deck is laid out. When the deck used for matching is laid out in correspondence to the target deck, any of the cards of the matching deck may fall below any one of the cards of the target deck. This is illustrated in the array below, where the columns correspond to the cards of the target deck and the entries in

R_1	R_2	R_3	R_4	B_1	B_2	B_3	B_4
R_1	R_1	R_1	R_1	R_1	R_1	R_1	R_1
R_2	R_2	R_2	R_2	R_2	R_2	R_2	R_2
R_3	R_3	R_3	R_3	R_3	R_3	R_3	R_3
R_4	R_4	R_4	R_4	R_4	R_4	R_4	R_4
B_1	B_1	B_1	B_1	B_1	B_1	B_1	B_1
B_2	B_2	B_2	B_2	B_2	B_2	B_2	B_2
B_3	B_3	B_3	B_3	B_3	B_3	B_3	B_3
B_4	B_4	B_4	B_4	B_4	B_4	B_4	B_4

Array I

R_1	R_2	R_3	R_4	B_1	B_2	B_3	B_4
x	x	x	x	1	1	1	1
x	x	x	x	1	1	1	1
x	x	x	x	1	1	1	1
x	x	x	x	1	1	1	1
1	1	1	1	x	x	x	x
1	1	1	1	x	x	x	x
1	1	1	1	x	x	x	x
1	1	1	1	x	x	x	x

Array II

the columns correspond to the cards of the matching deck. Now, whenever R_1 appears in a column labeled R_1, R_2, R_3, or R_4, this represents a matched pair; and likewise other reds in the red columns and the blacks in the black columns represent matched pairs. In every such case let us replace the symbol by x, and in all other cases by 1. In this way we form the second array.

Now, if we make all possible choices of one element from each column of the first array, with no two from any row, we have all of the 8! orders in which the cards of the matching deck may appear. Let us suppose that one of these is $R_1R_2B_1B_3R_3R_4B_2B_4$. This corresponds to $x, x, 1, 1, 1, 1, x, x$ in the second array, or, if we drop the

* T. E. Sterne, *The solution of a problem in probability*, Science, vol. 86, pp. 500–501, has calculated the first four moments for the case $s=5, t=5$; also note numerical results given by E. V. Huntington, *Exact probabilities in certain card-matching problems*, Science, vol. 86, pp. 499–500; and results given by C. E. Kellogg, *New evidence(?) for extra-sensory perception*, Scientific Monthly, vol. 45 (1937), pp. 331–341.

commas and multiply, it corresponds to x^4, where the exponent indicates that we have exactly four matched pairs. Forming every possible product from the second array, using only one element from each column and no two from any row, we have 8! terms of various degrees. Collecting like powers we have

$$\phi(x) = (4!)^2x^8 + 16(4!)^2x^6 + 36(4!)^2x^4 + 16(4!)^2x^2 + (4!)^2,$$

which means that, of all the possible arrangements, $(4!)^2$ produce eight matched pairs, $16(4!)^2$ produce six matched pairs, for example. If we divide the coefficients by 8! we have the probabilities. Thus Fry's problem, which requires the probability for six matched pairs, has the answer $16(4!)^2/8!$ or 8/35.

If we denote by $f(r)$ the frequency of r matched pairs, we may write the above polynomial in the form

$$\phi(x) = \sum_{r=0}^{8} f(r)x^r.$$

Then, obviously, $\phi(x)/8!$ is a generating function for factorial moments. Denoting the factorial moments by K, where

$$K_i = \frac{\sum_{r=0}^{8} r^{(i)}f(r)}{8!}, \qquad \text{if} \qquad r^{(i)} = r(r-1)\cdots(r-i+1),$$

we have

$$K_0 = \phi(1)/8!,\ K_1 = \phi'(1)/8!,\ \cdots,\ K_i = \phi^i(1)/8!.$$

The form of the function $\phi(x)$ needs detailed consideration. This polynomial was obtained from the array of x's and 1's by adding all the terms formed by choosing one element from each column but no two from any row. Such an array has been called a *permanent*,* and the symbol $+||+$ has been used for it. The properties of the permanent which will be useful in this paper are given in the next section.

3. **The permanent.** It is obvious that the permanent has properties analogous to those properties of the determinant which do not depend on the fact that the determinant vanishes when two rows (or two columns) are identical. For example, the permanent can be expanded in terms of minor permanents. Of particular importance is the fact that the derivative of a permanent in one variable is equal to the sum

* T. Muir, revised by W. H. Metzler, *A Treatise on the Theory of Determinants*, Muir, Albany, 1930, p. 19.

of the permanents formed by replacing, one column at a time, the elements of each column by their derivatives.

4. **Application of the permanent.** As noted above, we may identify $\phi(x)$ as a permanent, namely

$$\phi(x) = \begin{vmatrix} x & x & x & x & 1 & 1 & 1 & 1 \\ x & x & x & x & 1 & 1 & 1 & 1 \\ x & x & x & x & 1 & 1 & 1 & 1 \\ x & x & x & x & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & x & x & x & x \\ 1 & 1 & 1 & 1 & x & x & x & x \\ 1 & 1 & 1 & 1 & x & x & x & x \\ 1 & 1 & 1 & 1 & x & x & x & x \end{vmatrix}.$$

Obviously, $\phi(1) = 8!$. The first term of $\phi'(x)$ is

$$\begin{vmatrix} 1 & x & x & x & 1 & 1 & 1 & 1 \\ 1 & x & x & x & 1 & 1 & 1 & 1 \\ 1 & x & x & x & 1 & 1 & 1 & 1 \\ 1 & x & x & x & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & x & x & x & x \\ 0 & 1 & 1 & 1 & x & x & x & x \\ 0 & 1 & 1 & 1 & x & x & x & x \\ 0 & 1 & 1 & 1 & x & x & x & x \end{vmatrix} = 4 \cdot \begin{vmatrix} x & x & x & 1 & 1 & 1 & 1 \\ x & x & x & 1 & 1 & 1 & 1 \\ x & x & x & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & x & x & x & x \\ 1 & 1 & 1 & x & x & x & x \\ 1 & 1 & 1 & x & x & x & x \\ 1 & 1 & 1 & x & x & x & x \end{vmatrix},$$

and it is easy to see that each of the other seven terms is equivalent to this one. Therefore

$$\phi'(x) = 4 \cdot 8 \cdot \begin{vmatrix} x & x & x & 1 & 1 & 1 & 1 \\ x & x & x & 1 & 1 & 1 & 1 \\ x & x & x & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & x & x & x & x \\ 1 & 1 & 1 & x & x & x & x \\ 1 & 1 & 1 & x & x & x & x \\ 1 & 1 & 1 & x & x & x & x \end{vmatrix}, \quad \text{and} \quad \phi'(1) = 4 \cdot 8 \cdot 7! = 4 \cdot 8!.$$

In the permanent above, one of the blocks of x's has one less column and one less row than before. It is convenient, then, to denote

this permanent by $[1]$. Using the same system, we will denote by $[21]$ the permanent, where one block of x's is reduced by two rows and two columns and the other by one row and one column.

It can be readily verified that $\phi''(x)=4\cdot 8(3\cdot 3[2]+4\cdot 4[11])$ and $\phi''(1)=4\cdot 8(3^2+4^2)\cdot 6!=(4\cdot 5^2\cdot 8!)/7$. The succeeding derivatives may be written in similar fashion.

Using these results, we may compute the factorial moments and, from them, the moments about the mean.

5. **The solution of the general problem.** For the general problem we have t groups (instead of 2) of s like cards each (instead of 4). The general method of obtaining the moment-generating function, and from it the moments, is the same as was used for the special case, except that the work becomes somewhat more intricate and arduous. In the general case the function $\phi(x)$ is a permanent of n rows and n columns, $(st=n)$, and has t blocks of s x's each. Then

$$\begin{aligned}
\phi(x) &= [0],\\
\phi'(x) &= s^2t[1],\\
\phi''(x) &= (s^{(2)})^2t[2]+s^4t^{(2)}[1^2],\\
\phi'''(x) &= (s^{(3)})^2t[3]+3(s^{(2)}s)^2t^{(2)}[21]+s^6t^{(3)}[1^3],\\
\phi^{iv}(x) &= (s^{(4)})^2t[4]+4(s^{(3)}s)^2t^{(2)}[31]+3(s^{(2)})^4t^{(2)}[2^2]\\
&\quad +6(s^{(2)}s^2)^2t^{(3)}[21^2]+(s^4)^2t^{(4)}[1^4],
\end{aligned}$$

and, in general,

$$\phi^m(x)=\sum_\alpha C_\alpha\left\{(s^{(p_1)})^{\pi_1}(s^{(p_2)})^{\pi_2}\cdots(s^{(p_\nu)})^{\pi_\nu}\right\}^2 t^{(\rho_\alpha)}[p_1^{\pi_1}p_2^{\pi_2}\cdots p_\nu^{\pi_\nu}],$$

where the notation is explained as follows:

1. $p_i^{\pi_i}$ indicates that p_i appears π_i times in the partition of m.
2. The summation is to include all possible partitions of m.
3. For each term, $\sum\pi_i=\rho_\alpha$.
4. $C_\alpha=\dfrac{m!}{(p_1!)^{\pi_1}(p_2!)^{\pi_2}\cdots(p_\nu!)^{\pi_\nu}\pi_1!\pi_2!\cdots\pi_\nu!}$.

The work of Dwyer,* in a recent paper, proved to be of considerable assistance in obtaining a compact expression for the general derivative, as given above. It should be noted also that the numerical coefficients of the partitions which occur in our derivatives are given

* P. S. Dwyer, *Moments of any rational integral isobaric sample moment function*, Annals of Mathematical Statistics, vol. 8, pp. 21–65.

in his Table II.* For example, for the fourth derivative we might have used his "weight = 4" and read the proper numerical coefficients from the first column, namely, 1, 4, 3, 6, 1.

To evaluate these derivatives of $\phi(x)$ at $x=1$ it is sufficient to notice that, if the sum of the bracketed numbers is j, the permanent, represented by the bracket, has the value $(n-j)!$ for $x=1$. For example, each of the bracketed quantities in $\phi^{iv}(x)$ becomes $(n-4)!$ when x is set equal to 1.

The rest of the work necessary to calculate the moments about the mean is quite direct, although tedious. We obtain the factorial moments by dividing $\phi^i(x)$ by $n!$; then, by means of well known relations, we obtain the moments about the mean. Representing the moments about the mean by μ_i and replacing t by n/s, we obtain the mean and higher moments as follows:

$$\bar{r} = s,$$

$$\mu_2 = \frac{s(n-s)}{n-1},$$

$$\mu_3 = \frac{s(n-s)(n-2s)}{(n-1)(n-2)},$$

$$\mu_4 = \frac{s(n-s)}{(n-1)(n-2)(n-3)}\{(n-2s)(n-3s)(3s+1) + (s-1)(12ns - n - 18s^2 - 6s)\}.$$

Comparisons with the articles cited above indicate complete agreement between the results given for various special cases and the values obtained by the use of the formulas derived in this paper.

In conclusion, it should be mentioned that this general method can be applied to the more comprehensive case where the numbers in the t sets of the matching deck are not necessarily the same but may differ from set to set. Curiously, the mean is still s, but the variance is decreased.

CARNEGIE INSTITUTE OF TECHNOLOGY

* Loc. cit., pp. 30–32.

A THEOREM ON INTERIOR TRANSFORMATIONS*

G. T. WHYBURN

In an earlier paper† it has been shown that if $T(A)=B$ is a light interior transformation, that is, a continuous transformation mapping open sets into open sets and mapping no continuum into a single point, and if A is compact, then for every simple arc pq in B and any $p_0 \epsilon T^{-1}(p)$ there exists a simple arc p_0q_0 in A which maps topologically onto pq under T.

In this paper the following extension to arbitrary dendrites (that is, locally connected continua containing no simple closed curves) in B will be made.

THEOREM. *Let $T(A)=B$ be interior and light, where A is compact. For any dendrite D in B and any $x_0 \epsilon T^{-1}(D)$ there exists a dendrite E in A containing x_0 which maps topologically onto D under T.*

PROOF. Since the transformation $TT^{-1}(D)=D$ is interior,‡ clearly there is no loss of generality in assuming that $D=B$. Now let $H=\sum p_iq_i$ be an arc-development of B; that is, for each $i>0$, $p_iq_i \sum_{j=1}^{j=i-1} p_jq_j = q_i$, and p_i, $(i=1, 2, \cdots)$, and all points of $\overline{H}-H$ are end points of B, where we may suppose that $p_0=T(x_0)$. Now by the result quoted above, there exists an arc x_0y_0 such that $T(x_0y_0)=p_0q_0$ and T is topological on x_0y_0. Likewise there exists an arc x_1y_1 in A such that $T(x_1y_1)=p_1q_1$, T is topological on x_1y_1, and furthermore so that $y_1=T^{-1}(q_1)\cdot x_0y_0$. Similarly there is an arc x_2y_2 contained in A so that $T(x_2y_2)=p_2q_2$, T is topological on x_2y_2, and $y_2=T^{-1}(q_2)\cdot\sum_0^1 x_iy_i$. Continuing this process indefinitely we obtain $K=\sum x_iy_i$, so that $T(K)=H$, and T is topological on K. Clearly $\overline{K}$ is a continuum and $T(\overline{K})=B$. Hence our proof will be complete as soon as we show that T is 1 to 1 on $\overline{K}$, or, what amounts to the same thing, that for each $p \epsilon B$, $T^{-1}(p)\cdot\overline{K}$ reduces to a single point.

Now there exists a monotone decreasing sequence of connected neighborhoods V_1, V_2, V_3, $\cdots$ of p in B with $\delta(V_i)\to 0$. Furthermore,

---

* Presented to the Society, December 29, 1937, under the title *Interior transformations on certain curves.*

† See my paper in the Duke Mathematical Journal, vol. 3 (1937), p. 377, Theorem 4.1. Compare with Stoilow, Annales Scientifiques de l'École Normale Supérieure, vol. 63 (1928), pp. 347–382; and Montgomery, Transactions of this Society, vol. 42 (1937), pp. 328–329.

‡ See my paper, loc. cit., p. 370, Lemma 1.2.

since $\overline{H}-H$ contains only end points of B, it follows that $H\cdot V_1$, $H\cdot V_2$, $\cdots$ are connected sets which, of course, are open in H. Hence $K\cdot T^{-1}(V_1)$, $K\cdot T^{-1}(V_2)$, $\cdots$ is a monotone decreasing sequence of connected sets in K which are open in K. Then let $L=\lim K\cdot T^{-1}(V_i)$. Since $V_i\to p$, we have $L\subset T^{-1}(p)$. But, since L is necessarily connected and $T^{-1}(p)$ is totally disconnected, L must reduce to a single point $q\epsilon T^{-1}(p)$. Hence $T^{-1}(p)\cdot\overline{K}=q$, and our theorem is established.

That a similar conclusion cannot be obtained, if we permit D to contain simple closed curves, is readily seen, since it fails to hold even in the simple transformation $w=z^2$ of the circle $|z|=1$ into the circle $|w|=1$. Still more striking, however, is the following:

EXAMPLE. *There exist two connected graphs A and B and an interior transformation $T(A)=B$, where A contains no subset homeomorphic with B.*

Let $J=a_1b_1f_1c_1a_2b_2f_2c_2$ be a simple closed curve in a plane, where the points a_1, b_1, $\cdots$ are cyclically ordered on J as indicated. Let $a_1d_1f_2$ and $b_1e_1c_1$ be disjoint arcs lying within J except for their end points, and let d_1e_1 be an arc within J having only its end points in common with $a_1d_1f_2+b_1e_1c_1$. Similarly, let $f_1d_2a_2$ and $b_2e_2c_2$ be disjoint arcs lying, except for their end points, without J, and let d_2e_2 be an arc without J having just its end points in common with $f_1d_2a_2+b_2e_2c_2$. Finally let A be the graph thus constructed, that is,

$$\begin{aligned}A &= J + a_1d_1f_2 + b_1e_1c_1 + d_1e_1 + f_1d_2a_2 + b_2e_2c_2 + d_2e_2\\ &= a_1b_1 + b_1f_1 + f_1c_1 + c_1a_2 + b_1e_1 + e_1c_1 + a_1d_1 + d_1f_2 + d_1e_1\\ &\quad + a_2b_2 + b_2f_2 + f_2c_2 + c_2a_1 + b_2e_2 + e_2c_2 + a_2d_2 + d_2f_1 + d_2e_2.\end{aligned}$$

Then A is a graph of eighteen edges as indicated in the latter sum.

We now construct a graph B in 3-space as follows: Let θ be a θ-curve $bac+bec+bfc$ in a plane π. Let d be a point not in π, and let da, de, and df be arcs having just d in common and having just a, e, and f, respectively, in common with θ.

If we now let

$$\begin{aligned}B &= \theta + da + de + df\\ &= ab + bf + fc + ca + be + ec + ad + df + de,\end{aligned}$$

then B is a graph of nine edges as here indicated.

Now let $T(A)=B$ be the transformation which maps a_1b_1, a_2b_2 topologically into ab, b_1f_1, and b_2f_2 topologically into bf, and so on. In general T maps any edge x_iy_j of A topologically, preserving end

points, into the edge xy of B. Then T is an interior transformation. In fact T is a local homeomorphism which is 2 to 1.

Since A is a planar graph, whereas B is non-planar (B is, in fact, one of the two well known Kuratowski primitive skew curves), clearly A contains no subset homeomorphic with B. Incidentally this example shows that *planarity is not an interior property* (that is, it is not invariant under interior transformations).

UNIVERSITY OF VIRGINIA

ON THE TRANSFORMATION GROUP FOR DIABOLIC MAGIC SQUARES OF ORDER FOUR*

BARKLEY ROSSER AND R. J. WALKER

This paper concerns only magic squares of order four, and all statements of the paper are to be construed as applying only to magic squares of order four.

One says that

(1)

a	b	c	d
e	f	g	h
i	j	k	l
m	n	o	p

is a diabolic (or pan-diagonal or Nasik) magic square if $a, b, \cdots, p$ are $1, 2, \cdots, 16$ in some order, and each row, column, and diagonal adds up to 34. This is to include broken diagonals such as i, f, c, p, or c, h, i, n. A diabolic magic square clearly remains diabolic if subjected to the following transformations:

A. Reflection about the a, f, k, p diagonal.
B. Rotation through 90° counter-clockwise.
C. Putting the first column last.
D. Putting the first row last.

For many purposes it is convenient to consider a diabolic magic

---

* Presented to the Society, December 30, 1937.

square as mapped on a torus. That is, the square is bent and stretched until the top and bottom edges coincide and the right and left edges coincide. Then a diagonal like b, g, l, m is continuous and not broken.

THEOREM 1. *The four elements of any square of order two of a diabolic magic square of order four mapped on a torus add up to* 34.

This includes squares like a, d, m, p or e, h, i, l.

PROOF. Add together the two main diagonals and the second and third rows and subtract the first and fourth columns. This gives $f+g+j+k=34$. By use of transformations C and D, this proof can be applied to any square of order two.

THEOREM 2. *If a diabolic magic square of order four be mapped on a torus, then the sum of any element and the element which is two distant from it along a diagonal is* 17.

Note that if one starts at any element and moves two elements along any diagonal, one always stops at the same element regardless of the direction of the diagonal. Two elements so situated will here be called antipodal.

PROOF. Add together the first and the third rows, then add the first and third columns and the diagonals a, f, k, p and a, n, k, h; and subtract the diagonals e, b, o, l; i, f, c, p; m, j, g, d; and i, n, c, h. This gives $a+k=17$. By use of transformations C and D, this proof can be applied to any pair of antipodal elements.

By use of these two theorems, one can readily see that a diabolic magic square remains diabolic if subjected to the following transformation:

E. The transformation of (1) into

a	d	h	e
b	c	g	f
n	o	k	j
m	p	l	i

THEOREM 3. *All diabolic magic squares of order four can be derived from a single one by successive applications of* $A, B, C, D,$ *and* E.

PROOF. By use of C and D, 1 can be brought into the upper left-hand corner. Thus there is no loss of generality in taking $a=1$, $k=16$. Then c and d can at most be 14 and 15, so that $c+d\leqq 29$. Hence $b\geqq 4$. Similarly $c\geqq 4$, $d\geqq 4$, $e\geqq 4$, $i\geqq 4$, $m\geqq 4$. Since $a+b+e+f=34$, $a+d+e+h=34$, $a+b+m+n=34$, and $a+d+m+p=34$ by Theorem 1, one can similarly get $f\geqq 4$, $h\geqq 4$, $n\geqq 4$, $p\geqq 4$. Thus 2 and 3 can only occur in the positions g, j, l, and o. If 2 occurs in the o or l position, it can be brought to the j position by one or two applications of E. If 2 occurs in the j position, it can be brought to the g position by an application of A. So one can take $g=2$, $m=15$ with no loss of generality. If 3 occurs in the l or j position, it can be brought to the o position by one or two applications of E. Hence one takes $o=3$, $e=14$, $i=4$, $c=13$. Now $d=20-b$, and by Theorem 2, $l=17-b$, $j=b-3$. By Theorem 1, $f=19-b$, so that $h=b-1$, $n=18-b$, $p=b-2$. Hence we have to find values of b such that $b-3$, $b-2$, $b-1$, b, $17-b$, $18-b$, $19-b$, and $20-b$ are 5, 6, 7, 8, 9, 10, 11, and 12 in some order. Clearly $b=8$ or $b=12$. With $b=8$, we get

1	8	13	12
14	11	2	7
4	5	16	9
15	10	3	6

By applying C, B, and A successively to this we get the square that would result from putting $b=12$. Thus all diabolic magic squares can be obtained from the one shown by successive applications of A, B, C, D, and E.

THEOREM 4. *There are* 384 *diabolic magic squares of order four.*

PROOF. If X and Y are two transformations, denote by XY the transformation effected by first performing Y and then performing X. Then (a) $BA=AB^3$; (b) $CA=AD$; (c) $DA=AC$; (d) $CB=BD$; (e) $DB=BC^3$; (f) $DC=CD$; (g) $EA=AB^2CDE$; (h) $E^2A=B^2CDE^2$; (i) $EB=B^3CDE$; (j) $E^2B=ABCDE^2$; (k) $EC=C^3E^2$; (l) $E^2C=B^2CD^2E$; (m) $ED=ACE^2$; (n) $E^2D=AB^2DE$; (o) $A^2=B^4=C^4=D^4=E^3=$ the identical transformation. These relations were obtained by inspection and can be easily verified. By means of (a)–(o) one can, in any product of A, B, C, D, E, get all the E's to the right,

then all A's to the left, then all B's next to the A's, and finally all C's next to the B's; so that any product of A, B, C, D, E is equal to a product of the form $A^\alpha B^\beta C^\gamma D^\delta E^\epsilon$. Moreover A, B, C, and D all carry rows into rows and columns into columns, or rows into columns and columns into rows, so that neither E nor E^2 is a product of A, B, C, and D. Moreover, if a square is mapped on a torus, B, C, and D all leave the orientation unchanged, so that A is not a product of B, C, and D. Clearly B, C, and D are independent. Hence $A^\alpha B^\beta C^\gamma D^\delta E^\epsilon = A^a B^b C^c D^d E^e$ if and only if $\alpha \equiv a$ (2), $\beta \equiv b$ (4), $\gamma \equiv c$ (4), $\delta \equiv d$ (4), and $\epsilon \equiv e$ (3). Hence the order of the group generated by A, B, C, D, and E is $2 \times 4 \times 4 \times 4 \times 3 = 384$.* As each transformation is a permutation of the sixteen elements of the square, different transformations cannot yield the same square. So there are 384 diabolic magic squares.

This theorem disagrees with the result of D. N. Lehmer, *A census of squares of order* 4, *magic in rows, columns, and diagonals*, this Bulletin, vol. 39 (1933), p. 981. Lehmer's count was 48. It is easy to see the source of Lehmer's error. Lehmer found all squares which were normalized with respect to certain transformations, R, S, T, and U, and picked out the diabolic ones. However, Lehmer's transformation U destroys diabolism, thus certain of his normalized squares which are not diabolic give rise to diabolic squares when operated on by U. Hence Lehmer missed these squares in his count.

The present count was checked against Frenicle's list of magic squares. In so doing, we discovered the following startling result:

If there are two antipodal elements of a magic square of order four which add up to 17, then the magic square is diabolic.

The defining conditions of a magic square plus the single condition that a particular pair of antipodal elements add up to 17 give ten independent equations in the sixteen elements of the square, as contrasted with twelve independent equations arising from the defining conditions for a diabolic magic square. That the two sets of equations are equivalent is due to the diophantine condition that the elements must be $1, 2, \cdots, 16$ in some order.

* H. S. M. Coxeter has suggested that this group may be isomorphic with the symmetry group of the Cartesian frame in four dimensions, as the two groups are of the same order. That this is so can be shown as follows. The elements $R_1 = ABCE$, $R_2 = AB^3D$, $R_3 = BCE$, $R_4 = AB$, will generate our group, since $A = R_1R_2$, $B = R_1R_3R_4$, $C = R_4R_3R_1R_2R_1R_2$, $D = R_2R_4R_3R_2$, $E = R_1R_2R_2R_3$, as may be verified by the relations (a)–(o). The R_i satisfy the relations $R_4^2 = (R_1R_3)^2 = (R_1R_4)^2 = (R_2R_4)^2 = (R_1R_2)^3 = (R_2R_3)^3 = (R_3R_4)^4 = 1$, hence our group is a factor group of the above mentioned symmetry group (Coxeter, Journal of the London Mathematical Society, vol. 10 (1935), pp. 21–25). The group in question is designated by $[3^2, 4]$. Since the two groups have the same order they are isomorphic.

If one weakens the requirements still further and only asks that a square be magic in the rows and columns, then a pair of antipodal elements can add up to 17 without the square being diabolic. This is illustrated by

1	15	4	14
12	2	9	11
8	7	16	3
13	10	5	6

which is magic in rows and columns, but not in diagonals, and which has $a+k=e+o=17$.

An analogous treatment of the problem of finding all diabolic magic squares is given by Kraitchik on page 167 of his book, *La Mathématique des Jeux*, where he shows that all diabolic magic squares can be derived by successive applications of A, B, C, and D from three particular ones which he gives.

CORNELL UNIVERSITY

A NOTE ON REGULAR BANACH SPACES*

B. J. PETTIS

Introduction. For an element x of a Banach space B_0† it is well known that the functional

$$X_x(f) = f(x)$$

defined over $B_1=\overline{B}_0$, the Banach space composed of all linear functionals (real-valued additive and continuous functions) defined over B_0, is linear; moreover‡

$$\|X_x\|_{\overline{B}_1} = \|x\|_{B_0};$$

hence the additive operation $X_x=T(x)$ from B_0 to $B_2=\overline{B}_1$ is continuous and norm-preserving. In B_2 let $B_2^{(0)}$ denote the set of image ele-

* Presented to the Society, October 30, 1937.

† S. Banach, *Théorie des Opérations Linéaires*, Warsaw, 1932, p. 53. We shall use Banach's terminology.

‡ Banach, loc. cit., pp. 188–189.

ments of T. A space B_0 is *regular** if and only if

$$B_2 = B_2^{(0)}. \tag{0.1}$$

Examples of such spaces are L^p, l^p, $(1 < p < \infty)$, and the hyper-Hilbert spaces.

The purpose of the present note is to list in §1 some conditions necessary and sufficient for regularity, to give the form of the general linear operation from a regular space to the space l^1 of absolutely convergent series (§2), and to generalize some theorems already known concerning the differentiability of functions of bounded variation defined to a Banach space (§3). A function $Y(R)$, from the elementary figures in a euclidean figure R_0 to a B space, is of *bounded variation*, or BV, if and only if $\sum_{i=1}^{K} \| Y(R_i) \|$ is bounded over all finite sets of non-overlapping elementary figures $R_1, \cdots, R_K$ contained in R_0. The l.u.b. of such sums will be denoted by $\text{var}(Y; R_0)$. The last section includes the result (Theorem 7), analogous to that of J. A. Clarkson for uniformly convex spaces and those of Dunford and Morse,† that every function BV, from a linear interval to a regular B-space, is strongly differentiable‡ a.e. (almost everywhere).§

1. **Necessary and sufficient conditions for regularity.** We prove the following theorem:

THEOREM 1. *Each of the following conditions is necessary and sufficient that B_0 be regular*:

(1.1) *B_1 is regular*;

(1.2) *for linear sets in B_2 closure is equivalent to regular closure*;

(1.3) *B_0 is weakly complete, and for linear sets in B_2 regular closure is equivalent to weak closure as a set of functionals over B_1*;

(1.4) *$B_2^{(0)}$ is of the second category in B_2*;

(1.5) *each linear functional over $B_2^{(0)}$ is uniquely extensible over the whole of B_2.*

(0.1) implies (1.1). The operation $f_F = \overline{T}(F)$, adjoint (or associate)

---

* H. Hahn, *Über lineare Gleichungssysteme in linearen Räumen*, Journal für die Reine und Angewandte Mathematik, vol. 157 (1927), pp. 214–229.

† J. A. Clarkson, *Uniformly convex spaces*, Transactions of this Society, vol. 40 (1936), pp. 396–414; N. Dunford and A. P. Morse, *Remarks on the preceding paper of James A. Clarkson*, ibid., pp. 415–420.

‡ The definition of strong and weak differentiability will be given in §3.

§ See also Gelfand, *Zur Theorie abstrakter Funktionen*, Comptes Rendus de l'Académie des Sciences de l'URSS, vol. 17 (1937), Theorem 3. This theorem of Gelfand's was called to the writer's attention after the present paper had been accepted for publication.

to T and from $B_3=\overline{B}_2$ to B_1, assigns to each $F(X)$ the functional $f_F(x)$ defined by

$$f_F(x) \equiv F(T(x)) \equiv F(X_x);$$

hence for every X_x in $B_2^{(0)}$

$$F(X_x) = f_F(x) = X_x(f_F).$$

Since $B_2^{(0)}=B_2$, this implies for every F in B_3 an f_F in B_1 such that $F(X)=X(f_F)$ for all X in B_2; thus B_1 is regular.

(1.1) implies (1.2). If M is a closed linear set in B_2, then for X_0 not in M there exists an F_0 in B_3 such that $F_0(X_0)=1$, and $F_0(X)=0$ for X in M. Since B_1 is regular, there is an f_0 in B_1 such that $F_0(X)=X(f_0)$ for all X in B_2. Hence $X_0(f_0)=1$, $X(f_0)=0$ for X in M, and M is a regularly closed set of functionals over B_1.

(1.2) implies (1.3). In general, regular closure implies weak closure as a set of functionals, and this, in turn, is stronger than ordinary closure. From (1.2) we then have closure, weak closure as a set of functionals, and regular closure, all equivalent for linear sets in B_2. Moreover, the closed linear subspace $B_2^{(0)}$ is therefore weakly closed as a set of functionals over B_1; hence B_0 is weakly complete.

(1.3) implies (0.1). From the weak completeness of B_0 it follows easily that $B_2^{(0)}$ is weakly closed as a set of functionals. From the remainder of (1.3) the linear set $B_2^{(0)}$ is regularly closed. Hence if there existed an X_0 in B_2 that was not in $B_2^{(0)}$, there would then be an f_0 such that $X_0(f_0)=1$, while $X(f_0)=0$ for all X in $B_2^{(0)}$. But then we would have $f_0(x)=0$ for all x in B_0, that is, f_0 would be the identically zero functional over B_0, contrary to $X_0(f_0)=1$.

That (1.4) is equivalent to (0.1) follows from the fact that the set of image points of a linear operation from one Banach space to another is either of the first category in the contradomain space of the operation or consists of the whole of that space.

If B_2 contains an X_0 not in $B_2^{(0)}$, then, since $B_2^{(0)}$ is linear and closed, there is an F_0 in B_3 such that $F_0(X_0)=1$, and $F_0(X)=0$ for X in $B_2^{(0)}$. Thus the zero functional over $B_2^{(0)}$ is extensible in at least two different ways, which contradicts (1.5).

THEOREM 2. *If B_0 is separable, then each of the following conditions is necessary and sufficient that B_0 be regular:*

(1.6)* *B_0 is weakly complete and weakly compact;*

* (1.6) is well known; for the sufficiency see Banach, loc. cit., p. 189, and for the necessity, T. H. Hildebrandt, *Linear functional transformations in general spaces*, this Bulletin, vol. 37 (1931), p. 200.

(1.7) *B_0 is weakly complete and B_1 is separable*;

(1.8) *in B_1 weak convergence as a set of functionals is equivalent to weak convergence to an element as a set of elements.*

If B_0 is separable, so is $B_2^{(0)}$; hence if B_0 is in addition regular, then B_2, and therefore B_1, are separable. Thus from (1.3) the regularity and separability of B_0 together imply (1.7).

Condition (1.6) follows from (1.7). If $\{x_n\}$ is a bounded sequence in B_0, then, by diagonalizing, a subsequence $\{x_n'\}$ can be found such that $\lim_n f_i(x_n')$ exists for every f_i in a denumerable set dense in B_1. The bounded sequence $\{x_n'\}$ is then weakly convergent in B_0.

Thus if B_0 is separable, conditions (0.1), (1.6), and (1.7) are all equivalent.

The necessity of (1.8) is clear. If $\lim_n f_n(x)$ exists for every x, then $f_0(x) = \lim_n f_n(x)$ is a linear functional over B_0, and if B_0 is regular, then $F(f_n)$ converges to $F(f_0)$ for every F in B_2. If (1.8) holds and $F(f)$ is linear over B_1, then the convergence of $\{f_n\}$ to f_0, as a sequence of functionals over B_0, implies $\lim F(f_n) = F(f_0)$; hence $F(f)$ is weakly continuous over B_1. Since B_0 is separable, this insures the inclusion of F in $B_2^{(0)}$.†

THEOREM 3. *If B_0 is regular, so is every closed linear subspace B_0'.*

Suppose $X'(f')$ is a linear functional over the space B_1' conjugate to B_0'. Define $X(f) = X'(f')$, where f varies over B_1, and f' is the linear functional over B_0' determined by f. Since B_0 is regular, there exists an x in B_0 such that $X(f) = f(x)$ for every f in B_1. We have only to show that x is in B_0'; for then

$$X'(f') = X(f) = f(x) = f'(x)$$

for all f and hence for all f', since any f' can be extended to form an f. If x is not in the closed linear subset B_0' of B_0, there exists an f^* in B_1 such that $f^*(x) \neq 0$, $f^*(x') = 0$ for all x' in B_0'. But then $X(f^*) = f^*(x) \neq 0$, while $X(f^*) = X'(f^{*\prime}) = X'(0) = 0$, a contradiction that completes the proof.

If B_0 is regular, then any separable closed linear subspace is also regular and is therefore weakly complete and weakly compact. The space B_0 must then have these two properties also. This gives us the following corollary:

COROLLARY. *A regular space is weakly complete and weakly compact.*

2. **Operations from a regular space to the space l^1.** We prove the following theorem:

† Banach, loc. cit., p. 131.

THEOREM 4. *If B_0 is regular and the operation*

$$T(f) = \{f^i\}$$

from B_1 to l^1 (space of absolutely convergent series) is linear, then T is completely continuous and is defined by a series $\sum_1^\infty x_n$, unconditionally convergent† in B_0; conversely, such a series defines a linear completely continuous operation from B_1 to l^1. The norm of T is l.u.b.$_{\|f\|=1}\sum_1^\infty |f(x_n)|$.

Given a linear T, let $F=\overline{T}(m)$ be the adjoint linear operation from $\overline{l^1}=l^\infty$, the space of bounded sequences, to B_2. If $F_i=T(m_i)$, where m_i is the linear functional over l^1 defined by the sequence having 1 in the ith place and 0's elsewhere, then

$$F_i(f) = m_i(T(f)) = m_i(\{f^i\}) = f^i.$$

Since B_0 is regular, there exists an x_i such that

$$f^i = F_i(f) = f(x_i)$$

for all f in B_1. Hence

$$T(f) = \{f(x_i)\}. \tag{2.1}$$

From the weak completeness of B_0 and the fact that $\sum_{i=1}^\infty |f(x_i)| < \infty$ for all f, the series $\sum x_i$ is unconditionally convergent;‡ moreover, given $\epsilon>0$ there is an N_ϵ such that§

$$\sum_{N_\epsilon+1}^{\infty} |f(x_i)| < \epsilon \tag{2.2}$$

for all f with $\|f\| \leqq 2$. If $\|f_i^*\| \leqq 1$, then by diagonalizing we can find a subsequence $\{f_i\}$ such that $\lim_i f_i(x_n)$ exists for every n. This fact, combined with (2.2), gives

$$\sum_{n=1}^{\infty} |f_i(x_n) - f_j(x_n)| < 2\epsilon$$

for i, j greater than a suitably chosen K_ϵ. Thus $T(f)$ is completely continuous.

Conversely, suppose $\sum x_n$ converges unconditionally in B_0. The operation $T(f)$ in (2.1) is then defined and additive from B_1 to l^1. Since

† A series is unconditionally convergent if every reordering is convergent; W. Orlicz, *Beiträge zur Theorie der Orthogonalentwicklungen*, II, Studia Mathematica, vol. 1 (1929), pp. 241–255; p. 242. For a statement of equivalent definitions see Banach, loc. cit., p. 240.

‡ Orlicz, loc. cit., Theorem 2.

§ Orlicz, loc. cit., p. 246.

it is the "pointwise" limit of the linear operations

$$(2.3) \qquad T_n(f) = \{f(x_1), \cdots, f(x_n), 0, 0, \cdots\}, \qquad n = 1, 2, \cdots,$$

T must also be linear. The complete continuity follows from the first part of the proof.

The statement concerning the norm of T is an immediate consequence of (2.1).

Without resorting to (2.2) the complete continuity of T could have been proved by using the following facts: (i) T is linear, (ii) B_1 is weakly complete and weakly compact, (iii) the property of converging weakly to an element is preserved under linear operations, (iv) in l^1 weak and strong convergence are equivalent.

In view of (1.1), Theorem 4 can be stated as follows:

THEOREM 4'. *If B_0 is regular, an operation*

$$T(x) = \{f_x^i\}$$

from B_0 to l^1 is linear if and only if it is defined by an unconditionally convergent series $\sum f_i$ in B_1. If it is linear, it is completely continuous and

$$\| T \| = \text{l.u.b.}_{\|x\|=1} \sum_1^\infty |f_i(x)|.$$

It is clear from the forms of their general linear functionals that the spaces L^p, l^p, $(1 < p < \infty)$, are regular. Applying Theorem 4' and (2.2) to these spaces we obtain the following corollary:

COROLLARY.† *An operation $T(x)$ from L^p (l^p), $(1 < p < \infty)$, to l^1 is linear if and only if there exists in $L^{p'}$ $(l^{p'})$, $(p' = p/p-1)$, a sequence $\{y_i\} = \{y_i(s)\}$ $(\{y_i\} = \{\{y_i^j\}\})$ such that*

$$(1) \qquad \sum_{i=1}^\infty \left| \int_0^1 y_i(s)x(s)ds \right| < \infty \quad \left(\sum_{i=1}^\infty \left| \sum_{j=1}^\infty y_i^j x^j \right| < \infty \right)$$

for every $x = x(s)$ $(x = \{x^j\})$ in L^p (l^p); that is, $\sum_1^\infty y_i$ converges unconditionally in $L^{p'}$ $(l^{p'})$, and

$$(2) \qquad T(x) = \left\{ \int_0^1 y_i(s)x(s)ds \right\} \quad \left(T(x) = \left\{ \sum_j y_i^j x^j \right\} \right).$$

If T is linear, then

† In this corollary the case l^p to l^1 is already known. See H. R. Pitt, *A note on bilinear forms*, Journal of the London Mathematical Society, vol. 11 (1936), pp. 174–180; and L. W. Cohen and N. Dunford, *Transformations on sequence spaces*, Duke Mathematical Journal, vol. 3 (1937), pp. 689–701.

(i) *T is completely continuous;*

(ii) *given $\epsilon>0$ there exists an N_ϵ such that*

$$\sum_{i=N_\epsilon}^{\infty}\left|\int_0^1 y_i(s)x(s)ds\right|<\epsilon \qquad \left(\sum_{i=N_\epsilon}^{\infty}\left|\sum_{j=1}^{\infty}y_i{}^jx^j\right|<\epsilon\right)$$

uniformly for all x with

$$\int_0^1|x(s)|^p ds\leqq 1 \quad \left(\sum_{j=1}^{\infty}|x^j|^p\leqq 1\right);$$

(iii)
$$\|T\| = \underset{\int_0^1|x(s)|^p ds\leqq 1}{\text{l.u.b.}} \sum_{i=1}^{\infty}\left|\int_0^1 y_i(s)x(s)ds\right|$$

$$\left(\|T\| = \underset{\sum_{j=1}^{\infty}|x^j|^p\leqq 1}{\text{l.u.b.}} \sum_{i=1}^{\infty}\left|\sum_{j=1}^{\infty}y_i{}^jx^j\right|\right).$$

3. **Differentiation in regular spaces.** In this section we consider the differentiability of functions BV from a linear interval to a regular space. Our main purpose is to prove the theorem stated in the introduction. Every such function is strongly differentiable a.e. A function $Y(R)$, from the elementary figures in a euclidean figure R_0 to a Banach space B_0, is said to be *strongly differentiable* at a point s in the interior of R_0 if there exists an element $y(s)$ of B_0 such that for cubes I lying in R_0 and containing s it is true that $\|Y(I)/|I|-y(s)\|$ tends to 0 with $|I|$, the measure of I. The function $Y(R)$ is *weakly differentiable* at s if there exists a $y(s)$ such that

$$\lim_{\substack{|I|\to 0\\ s\in I}}|f(Y(I)/I)-f(y(s))|=0$$

for every linear functional f.

THEOREM 5. *If $Y(R)$ is defined and BV from the figures in a euclidean figure R_0 to a Banach space B_0, and if $Y(R)$ is weakly differentiable a.e. in R_0 to the function $y(s)$, then $y(s)$ is Bochner integrable.**

Let I_0 be a cube containing R_0, and $\{\Pi_n\}$ a sequence of partitions of I_0 into a finite number of non-overlapping (except on their boundaries) and non-degenerate subcubes, with Π_{n+1} a repartition of Π_n and $\lim_{n\to\infty}\text{norm }\Pi_n=0$. If s lies in the interior of a cube I_{in} of Π_n and R_0 contains I_{in}, then define $y_n(s)$ as $Y(I_{in})/|I_{in}|$; otherwise let $y_n(s)$ vanish. Then a.e. in R_0 we have $\{y_n(s)\}$ converging weakly to $y(s)$, hence

* S. Bochner, *Integration von Funktionen, deren Werte die Elemente eines Vektorräumes sind*, Fundamenta Mathematicae, vol. 20 (1933), pp. 262–276.

$$\|y(s)\| \leqq \liminf_n \|y_n(s)\|$$

a.e. in R_0. From Fatou's lemma and the inequality

$$\int_{R_0} \|y_n(s)\| ds \leqq \text{var } (Y; R_0) < \infty,$$

the real-valued $\liminf_n \|y_n(s)\|$ must be summable. The abstract function $y(s)$ is measurable in the sense of Bochner,* since a.e. in R_0 it is a weak derivative.† Hence $\|y(s)\|$ is measurable and summable, which implies that the measurable $y(s)$ is Bochner integrable.

COROLLARY.‡ *If, in addition to the assumptions in Theorem* 5, $Y(R)$ *is additive and AC (absolutely continuous), then*

$$\text{(A)} \qquad Y(R) = \int_R y(s) ds,$$

and $Y(R)$ *is strongly differentiable a.e. to the function* $y(s)$.

In view of the strong differentiability a.e. of the Bochner integral,§ it is sufficient to prove (A). Let $Z(R) = \int_R y(s) ds$; then $\| f(Z(R)) = \int_R f(y(s)) ds$ for every f in $B_1 = \overline{B}_0$ and every R in R_0. But $f(Y(R))$ is a real-valued additive and AC function that is differentiable a.e. to $f(y(s))$, and hence $f(Y(R)) = \int_R (d/ds f(Y(R))) ds = \int_R f(y(s)) ds$ for every f and R. Thus $Y(R) = Z(R)$ for every R.

Because of this corollary a theorem of Dunford and Morse¶ may be altered to read:

THEOREM 6. *A Banach space* B_0 *has the property* (DBV), *namely, that every BV function from a linear interval to* B_0 *is strongly differentiable a.e., and its derivative is Bochner integrable, if and only if every Lipschitzian function from a linear interval to* B_0 *is weakly differentiable a.e.*

THEOREM 7. *If* B_0 *is regular, then* B_0 *has property* (DBV).

---

* Bochner, loc. cit., p. 263.

† See *On integration in vector spaces*, a paper by the present author which is to appear in the Transactions of this Society.

‡ Theorem 5 and the corollary are generalizations of theorems due to Clarkson, loc. cit., pp. 409–410.

§ Bochner, loc. cit., p. 269.

‖ N. Dunford, *Integration and linear operations*, Transactions of this Society, vol. 40 (1936), pp. 474–494; p. 475, Theorem 2.3.

¶ Loc. cit., p. 415.

Suppose $Y(t)$ is defined and Lipschitzian from a linear interval $[a, b]$ to a regular space B_0. From the Lipschitzian property it is clear that every value of $Y(t)$ lies in the separable closed linear subspace B_0' generated by the values of $Y(t)$ at the rationals. By (1.7) and Theorem 3 there exists a denumerable set $\{f_i'\}$ dense in the space B_1' adjoint to B_0'; and B_0' is weakly complete. Let $Y(R)$ be the additive figure function determined by $Y(t)$. Then $f_i'(Y(R))$ is a real-valued additive Lipschitzian figure function; hence there exists in $[a, b]$ a measurable set S with $|S| = b - a$ and such that s in S implies $f_i'(Y(R))$ differentiable at s for each i. Thus if s is in S, and if $\{I_n\}$ are intervals closing down on s, it follows that

$$\lim_n \frac{f'(Y(I_n))}{|I_n|} = \lim_n f'\left(\frac{Y(I_n)}{|I_n|}\right)$$

exists for every f' in a set dense in B_1'; moveover, by the Lipschitzian property $\|Y(I_n)/|I_n|\| \leqq K$. These two conditions imply that $\lim f'(Y(I_n)/|I_n|)$ exists for every f' in B_1', that is, $\{Y(I_n)/|I_n|\}$ is a weakly convergent sequence. Since B_0' is weakly complete there is an element $y(s)$ to which this sequence converges weakly. If $\{I_n\}$ is another sequence of intervals closing down on s, then $\{Y(I_n)/|I_n|\}$ is also weakly convergent, and for every f'

$$\lim_n f'(Y(I_n)/|I_n|) = \lim_n f'(Y(I_n)/|I_n|) = f'(y(s)).$$

Thus $Y(R)$, as a function defined to B_0', is weakly differentiable a.e. in $[a, b]$. Since any f in B_1 defines an f' in B_1', $Y(R)$, as defined to B_0, has the same property. On application of Theorem 6, the proof is complete.

If L denotes the space of functions Lebesgue integrable on $[0, 1]$, we have the following corollary:

COROLLARY. *If B_0 is regular, then $T(\phi)$ from L to B_0 is linear if and only if there exists an essentially bounded and Bochner measurable function $x(s)$ on $[0, 1]$ to B_0 such that*

$$T(\phi) = \int_0^1 x(s)\phi(s)ds,$$

the integration being in the Bochner sense. The norm of T is ess. sup. $\|x(s)\|$.

To obtain this from the theorem the reader is referred to a proof by Dunford.*

THE UNIVERSITY OF VIRGINIA

* Loc. cit., p. 483, Theorem 5.3.

RELATIONS AMONG THE FUNDAMENTAL SOLUTIONS OF THE GENERALIZED HYPERGEOMETRIC EQUATION WHEN $p=q+1$. I. NON-LOGARITHMIC CASES*

F. C. SMITH

It is the purpose of this paper to develop the relations among the non-logarithmic solutions of the equation

$$(1) \qquad \left\{ \prod_{t=1}^{p} (\theta + a_t) - \frac{1}{z}\theta \prod_{t=1}^{q} (\theta + c_t - 1) \right\} y = 0, \quad p = q+1,$$

where $\theta = z(d/dz)$. The results of this paper generalize those of Mehlenbacher,† who studied the case in which $p=2$ and $q=1$, and extend those of Barnes‡ who obtained the asymptotic developments of the non-logarithmic solutions of (1) for those cases in which $p<q+1$.

The succeeding analysis will be simplified if we rewrite equation (1) in the equivalent form

$$(2) \qquad \left\{ \prod_{t=1}^{q+1} (\theta + a_t) - \frac{1}{z} \prod_{t=1}^{q+1} (\theta + c_t - 1) \right\} y = 0, \qquad c_{q+1} = 1.$$

If no two of the a_t or c_t are equal or differ by an integer, then equation (2) has $q+1$ linearly independent solutions about the point $z=0$, which may be written

$$(3) \qquad Y_{0j} = z^{1-c_j} \prod_{t=1}^{q+1} \frac{\Gamma(1 + c_t - c_j)}{\Gamma(1 + a_t - c_j)} \sum_{n=0}^{\infty} \prod_{t=1}^{q+1} \frac{\Gamma(1 + a_t - c_j + n)}{\Gamma(1 + c_t - c_j + n)} z^n,$$
$$j = 1, 2, 3, \cdots, q+1; \; c_{q+1} = 1; \; |z| < 1;$$

and $q+1$ linearly independent solutions about the point $z = \infty$ which may be written

$$(4) \qquad Y_{\infty j} = z^{-a_j} \prod_{t=1}^{q+1} \frac{\Gamma(1 - a_t + a_j)}{\Gamma(1 - c_t + a_j)} \sum_{n=0}^{\infty} \prod_{t=1}^{q+1} \frac{\Gamma(1 - c_t + a_j + n)}{\Gamma(1 - a_t + a_j + n)} \frac{1}{z^n},$$
$$j = 1, 2, 3, \cdots, q+1; \; c_{q+1} = 1; \; |z| > 1.$$

---

* Presented to the Society, December 30, 1937. The logarithmic cases will be considered in a later paper.

† See this Bulletin, Abstract 42-5-169.

‡ E. W. Barnes, *The asymptotic expansions of integral functions defined by generalized hypergeometric series*, Proceedings of the London Mathematical Society, (2), vol. 5 (1907), pp. 59–116.

In this paper, we shall develop the relations among solutions (3) and (4) of equation (2).

In this connection, we may state the following theorem:

THEOREM 1. *If no two of the a_t or c_t are equal or differ by an integer, then the solutions* (3) *of equation* (2), *when extended analytically outside their circle of convergence, may be expressed linearly in terms of solutions* (4) *in the following forms, it being understood throughout that* $0 < \arg z < 2\pi$:

$$Y_{0j} = \sum_{k=1}^{q+1} \left\{ e^{i\pi(a_k - c_j + 1)} \frac{\Gamma(c_j - a_k)\Gamma(1 + c_k - c_j)}{\Gamma(c_k - a_k)} \prod_{\substack{t=1 \\ t \neq k}}^{q+1} \frac{\Gamma(a_t - a_k)\Gamma(1 + c_t - c_j)}{\Gamma(c_t - a_k)\Gamma(1 + a_t - c_j)} Y_{\infty k} \right\}, \tag{5}$$

$$j = 1, 2, 3, \cdots, q + 1; \; c_{q+1} = 1.*$$

PROOF. It is well known that the function

$$g(w) = \prod_{t=1}^{q+1} \frac{\Gamma(1 + a_t - c_j + w)}{\Gamma(1 + c_t - c_j + w)} \tag{6}$$

satisfies the relation

$$g(w) = w^{\sum_{t=1}^{q+1}(a_t - c_t)}[1 + \gamma(w)], \qquad \lim_{|w| \to \infty} \gamma(w) = 0, \; |\arg w| < \pi.\dagger$$

Therefore,

$$|g(w)| \leqq |w^{\sum_{t=1}^{q+1}(a_t - c_t)}| \, [1 + |\gamma(w)|] = |w|^R e^{-I\phi}[1 + |\gamma(w)|],$$

where R and I denote the real and imaginary parts of $\sum_{t=1}^{q+1}(a_t - c_t)$, respectively, and where $\phi = \arg w$. For sufficiently large values of $|w|$, we have $|\gamma(w)| < \epsilon$, $|\arg w| < \pi$. Thus, for such values of $|w|$

$$|g(w)| < |w|^R e^{-I\phi}(1 + \epsilon) < |w|^R e^{I\pi}(1 + \epsilon) < K|w|^R,$$

where K is a constant independent of w. The function $g(w)$, therefore, satisfies the conditions of a general theorem due to Ford,‡ but, as ap-

* As far as relation (5) itself is concerned, the restriction $c_{q+1} = 1$ is unnecessary. The restriction is necessary, however, in order that the Y_{0j} and the $Y_{\infty j}$ be solutions of equation (2).

† See, for example, Milne-Thomson, *The Calculus of Finite Differences*, Macmillan, 1933, p. 255.

‡ See Ford, *The Asymptotic Developments of Functions Defined by Maclaurin Series*, University of Michigan Press, 1936, pp. 5, 9.

pears from (6), has simple poles at the points

$$(7)\quad w = c_j - a_k - n - 1,\quad n = 0, 1, 2, \cdots;\ k = 1, 2, 3, \cdots, q+1.$$

An application of this theorem to the function

$$(8)\qquad f(z) = \sum_{n=0}^{\infty} g(n)z^n = \sum_{n=0}^{\infty} \prod_{t=1}^{q+1} \frac{\Gamma(1 + a_t - c_j + n)}{\Gamma(1 + c_t - c_j + n)} z^n$$

gives the result

$$(9)\qquad f(z) \sim -\sum_{n=1}^{\infty} \prod_{t=1}^{q+1} \frac{\Gamma(1 + a_t - c_j - n)}{\Gamma(1 + c_t - c_j - n)} \frac{1}{z^n} - \sum_{k=1}^{q+1} \sum_{n=0}^{\infty} R_{n,k},$$
$$0 < \arg z < 2\pi,$$

where $R_{n,k}$ represents the residue of the function

$$(10)\qquad \frac{\pi(-z)^w g(w)}{\sin \pi w} = (-z)^w \Gamma(w)\Gamma(1-w) \prod_{t=1}^{q+1} \frac{\Gamma(1 + a_t - c_j + w)}{\Gamma(1 + c_t - c_j + w)}$$

at the point $w = c_j - a_k - n - 1$. But the first series on the right in (9) drops out since the denominator of every term contains a factor for which $t = j$, and

$$\frac{1}{\Gamma(1 + c_j - c_j - n)} = \frac{1}{\Gamma(1-n)} = 0, \qquad n = 1, 2, 3, \cdots.$$

It follows that our problem reduces to that of computing the residue term.

For this purpose, consider

$$(11)\quad R_{n,k} = \frac{1}{2\pi i} \int_{C_{n,k}} (-z)^w \Gamma(w)\Gamma(1-w) \prod_{t=1}^{q+1} \frac{\Gamma(1 + a_t - c_j + w)}{\Gamma(1 + c_t - c_j + w)}\, dw,$$

where $C_{n,k}$ surrounds the pole $w = c_j - a_k - n - 1$ of (10) but no other pole of (10). If in (11) we replace w by $-(w - c_j + a_k + n + 1)$, we obtain

$$R_{n,k} = -\frac{(-z)^{c_j - a_k - n - 1}}{2\pi i}$$
$$(12)\qquad \cdot \int_{C_j} (-z)^{-w} \frac{\Gamma(c_j - a_k - w - n - 1)\Gamma(w - c_j + a_k + n + 2)\Gamma(-w-n)}{\Gamma(c_k - a_k - w - n)}$$
$$\cdot \prod_{\substack{t=1 \\ t \neq k}}^{q+1} \frac{\Gamma(a_t - a_k - w - n)}{\Gamma(c_t - a_k - w - n)}\, dw,$$

where C_0 surrounds the origin. By several applications of the relation

$$\Gamma(1-w) = \frac{\pi}{\Gamma(w)\sin \pi w}, \tag{13}$$

we may change (12) to the form

$$\begin{aligned} R_{n,k} = &-\frac{(-z)^{c_j-a_k-n-1}(-1)^n}{2\pi i} \\ &\int_{C_0} \frac{\pi(-z)^{-w}}{\sin \pi w} \frac{\Gamma(c_j-a_k-w)\Gamma(w-c_j+a_k+1)}{\Gamma(c_k-a_k-w)\Gamma(w-c_k+a_k+1)} \\ &\cdot \prod_{\substack{t=1 \\ t\neq k}}^{q+1} \frac{\Gamma(a_t-a_k-w)\Gamma(w-a_t+a_k+1)}{\Gamma(c_t-a_k-w)\Gamma(w-c_t+a_k+1)} \\ &\cdot \prod_{t=1}^{q+1} \frac{\Gamma(w-c_t+a_k+n+1)}{\Gamma(w-a_t+a_k+n+1)}\, dw, \end{aligned} \tag{14}$$

which by virtue of a well known theorem in the calculus of residues* is equal to

$$\begin{aligned} R_{n,k} = &-\frac{(-z)^{c_j-a_k-1}}{z^n} \frac{\Gamma(c_j-a_k)\Gamma(1-c_j+a_k)}{\Gamma(c_k-a_k)\Gamma(1-c_k+a_k)} \\ &\cdot \prod_{\substack{t=1 \\ t\neq k}}^{q+1} \frac{\Gamma(a_t-a_k)\Gamma(1-a_t+a_k)}{\Gamma(c_t-a_k)\Gamma(1-c_t+a_k)} \cdot \prod_{t=1}^{q+1} \frac{\Gamma(1-c_t+a_k+n)}{\Gamma(1-a_t+a_k+n)}. \end{aligned} \tag{15}$$

Therefore,

$$\begin{aligned} \sum_{n=0}^{\infty} R_{n,k} = &-e^{i\pi(a_k-c_j+1)} z^{c_j-a_k-1} \frac{\Gamma(c_j-a_k)\Gamma(1-c_j+a_k)}{\Gamma(c_k-a_k)\Gamma(1-c_k+a_k)} \\ &\cdot \prod_{\substack{t=1 \\ t\neq k}}^{q+1} \frac{\Gamma(a_t-a_k)\Gamma(1-a_t+a_k)}{\Gamma(c_t-a_k)\Gamma(1-c_t+a_k)} \cdot \sum_{n=0}^{\infty} \prod_{t=1}^{q+1} \frac{\Gamma(1-c_t+a_k+n)1}{\Gamma(1-a_t+a_k+n)z^n} \\ = &-e^{i\pi(a_k-c_j+1)} z^{c_j-1} \frac{\Gamma(c_j-a_k)\Gamma(1-c_j+a_k)}{\Gamma(c_k-a_k)} \\ &\cdot \prod_{\substack{t=1 \\ t\neq k}}^{q+1} \frac{\Gamma(a_t-a_k)}{\Gamma(c_t-a_k)} Y_{\infty k}. \end{aligned} \tag{16}$$

Thus,

* See Ford, op. cit., p. 5, equation (20).

$$\sum_{k=1}^{q+1}\sum_{n=0}^{\infty} R_{n,k} = -\, z^{c_j-1}\sum_{k=1}^{q+1}\left\{e^{i\pi(a_k-c_j+1)}\frac{\Gamma(c_j-a_k)\Gamma(1-c_j+a_k)}{\Gamma(c_k-a_k)}\right.$$

$$\left.\cdot\prod_{\substack{t=1\\ t\neq k}}^{q+1}\frac{\Gamma(a_t-a_k)}{\Gamma(c_t-a_k)}\,Y_{\infty k}\right\}. \tag{17}$$

Substituting (17) into (9) and recalling that the first series of (9) is no longer present, we have

$$f(z) \sim z^{c_j-1}\sum_{k=1}^{q+1}\left\{e^{i\pi(a_k-c_j+1)}\frac{\Gamma(c_j-a_k)\Gamma(1-c_j+a_k)}{\Gamma(c_k-a_k)}\right.$$

$$\left.\cdot\prod_{\substack{t=1\\ t\neq k}}^{q+1}\frac{\Gamma(a_t-a_k)}{\Gamma(c_t-a_k)}\,Y_{\infty k}\right\}. \tag{18}$$

Since each $Y_{\infty k}$ which appears in (18) is defined by a series which converges for $|z|>1$, we may replace $\sim$ by $=$ in (18). Then, upon noting that

$$Y_{0j} = z^{1-c_j}\prod_{t=1}^{q+1}\frac{\Gamma(1+c_t-c_j)}{\Gamma(1+a_t-c_j)}f(z), \tag{19}$$

we see that (18) is equivalent to the desired result (5).

By means of a similar proof, we may establish the following theorem:

THEOREM 2. *If no two of the a_t or c_t are equal or differ by an integer, then the solutions* (4) *of equation* (2), *when extended analytically outside their circle of convergence, may be expressed linearly in terms of solutions* (3) *in the following forms, it being understood throughout that* $0<\arg\,(1/z)<2\pi$:

$$Y_{\infty j} = \sum_{k=1}^{q+1}\left\{e^{i\pi(a_j-c_k+1)}\frac{\Gamma(c_k-a_j)\Gamma(1-a_k+a_j)}{\Gamma(c_k-a_k)}\right.$$

$$\left.\cdot\prod_{\substack{t=1\\ t\neq k}}^{q+1}\frac{\Gamma(c_k-c_t)\Gamma(1-a_t+a_j)}{\Gamma(c_k-a_t)\Gamma(1-c_t+a_j)}\,Y_{0k}\right\}, \tag{20}$$

$$j = 1, 2, 3, \cdots, q+1;\ c_{q+1} = 1.$$

THE COLLEGE OF ST. FRANCIS

A NOTE ON FREDHOLM-STIELTJES INTEGRAL EQUATIONS*

F. G. DRESSEL

1. **Introduction.** The object of this paper is to show that the integral equation†

$$f(x) = m(x) + \lambda \int_0^1 f(y)dG(x, y), \qquad 0 \leqq x, y \leqq 1, \tag{1}$$

can be changed into an ordinary Fredholm equation when $G(x, y)$ is absolutely continuous $g(y)$.‡ The integration is carried out in the Young-Stieltjes sense, and $g(y)$ is a bounded, monotone increasing function.

2. **Lemmas.** If $h(x)$ is of bounded variation and we set $h(x) = h(0)$, $(x < 0)$, and $h(x) = h(1)$, $(x > 1)$, then we may define the completely additive function of sets $\bar{h}(e)$ by

$$\bar{h}(e) = h(x_2 + 0) - h(x_1 - 0), \qquad e = e(x_1 \leqq t \leqq x_2).$$

Using this notation we have the following lemma:

LEMMA 1. *If $f(x)$ is measurable Borel then*

$$\int_0^1 f(x)dh(x) = \int_0^1 f(x)d\bar{h},$$

the left side being Young-Stieltjes integration, the right Radon-Stieltjes.

In case one integral does not exist the equality sign is taken to mean that the other integration is non-existent. Because of the properties of the integrals under consideration, we need only prove the equality for the functions

$$f_1(x) = 1, \ x = \alpha, \qquad f_2(x) = 1, \ 0 \leqq \alpha < x < \beta \leqq 1,$$
$$= 0, \ x \neq \alpha; \qquad = 0, \text{ elsewhere.}$$

* Presented to the Society, December 29, 1936.

† For a discussion of (1) see G. C. Evans and O. Veblen, *The Cambridge Colloquium Lectures on Mathematics*, American Mathematical Society Colloquium Publications, vol. 5, 1922, p. 101.

‡ For terminology see Alfred J. Maria, *Generalized derivatives*, Annals of Mathematics, vol. 28 (1926–1927), pp. 419–432. I am much indebted to Mr. Maria for many valuable suggestions.

All functions used in the present paper are assumed to be measurable Borel.

We have

$$\int_0^1 f_1(x)dh(x) = h(\alpha+0) - h(\alpha-0) = \bar{h}(\alpha) = \int_0^1 f_1(x)d\bar{h};$$

$$\int_0^1 f_2(x)dh(x) = h(\beta-0) - h(\alpha+0) = \bar{h}(e) = \int_0^1 f_2(x)d\bar{h},$$

where e is the open set $\alpha<t<\beta$.*

LEMMA 2. *If $G(x)$ is absolutely continuous with respect to the bounded monotone increasing function $g(x)$, then*

$$\int_0^1 f(x)dG(x) = \int_0^1 f(x)DG(x)dg(x),$$

where $DG(x)$ is the derivative or one of the derived numbers of $G(x)$ with respect to $g(x)$.

Mr. Maria† has made the important step in the proof of the lemma by showing that

$$G(x_2+0) - G(x_1-0) = \int_E DG(x)d\bar{g},$$

where E is the set $x_1 \leqq t \leqq x_2$. For the function $f_1(x)$, making use of Lemma 1, we have

$$\int_0^1 f_1(x)dG(x) = G(\alpha+0) - G(\alpha-0),$$

$$\int_0^1 f_1(x)DG(x)dg(x) = \int_0^1 f_1(x)DG(x)d\bar{g} = \int_E DG(x)d\bar{g}$$
$$= G(\alpha+0) - G(\alpha-0),$$

where E is the point α. For $f_2(x)$ we have, if e is the open set $\alpha<x<\beta$,

$$\int_0^1 f_2(x)dG(x) = G(\beta-0) - G(\alpha+0),$$

$$\int_0^1 f_2(x)DG(x)dg(x) = \int_0^1 f_2(x)DG(x)d\bar{g} = \int_e DG(x)d\bar{g}$$
$$= G(\beta-0) - G(\alpha+0).$$

From the above material the lemma readily follows.

* The same reasoning shows that $\int_0^x f(t)dh(t)$ is equal to $\int_0^x f(t)d\bar{h}$, for $0<x\leqq 1$, if $h(t)$ is continuous from the right except perhaps at $x=0$.

† Loc. cit., p. 430.

3. **Transformations.** Our first theorem is the following.

THEOREM 1. *If $G(x, y)$ is absolutely continuous $g(y)$ then equation* (1) *can be written in the form*

$$f(x) = m(x) + \lambda \int_0^1 K(x, y) f(y) dg(y), \tag{2}$$

where $K(x, y) = DG(x, y)$, the derivative being taken with respect to $g(y)$, a bounded monotone increasing function.

This is immediate from Lemma 2.

THEOREM 2. *If $m(x)$ and $K(x, y)$ are bounded, then the solution of* (1) *and* (2), *except for characteristic values of λ, can be written*

$$f(x) = m(x) + \lambda \int_0^1 \frac{D(x, y; \lambda)}{D(\lambda)} m(y) dg(y), \tag{3}$$

where

$$D(\lambda) = 1 - \lambda \int_0^1 K(s, s) dg(s) + \cdots,$$

$$D(x, y; \lambda) = K(x, y) - \lambda \int_0^1 \begin{vmatrix} K(x, y) & K(x, s) \\ K(s, y) & K(s, s) \end{vmatrix} dg(s) + \cdots.$$

The proof follows along the same lines as in the ordinary case. We now state a corollary of Theorem 2 that represents most of the known results concerning solutions of equation (1).

COROLLARY.* *If $|G(x, y_2) - G(x, y_1)| \leqq |g(y_2) - g(y_1)|$, then, excepting characteristic values, equation* (1) *has* (3) *as a solution.*

Any result for the ordinary Fredholm equation carries a related result for equation (1). To see this, we assume without loss of generality that $g(y_1) < g(y_2)$ if $y_1 < y_2$, and apply to (2) the transformation†

$$\beta(s) = \limsup E_y(s \geqq g(y)), \qquad g(0) \leqq s \leqq g(1),$$

$$f(x) = m(x) + \lambda \int_0^1 K(x, y) f(y) dg(y) \tag{4}$$

$$= m(x) + \lambda \int_{g(0)}^{g(1)} K(x, \beta(s)) f(\beta(s)) ds.$$

* This includes the case handled by W. C. Randels, *On Volterra-Stieltjes integral equations*, Duke Mathematical Journal, vol. 1 (1935), pp. 538–542.

† Banach, *Théorie des Opérations Linéaires*, Warsaw, 1932, p. 6.

If we let ω be any of the possible solutions of

$$x = \beta(\omega), \qquad g(0) \leqq \omega \leqq g(1),$$

we may write (4) in the form

$$F(\omega) = M(\omega) + \lambda \int_{g(0)}^{g(1)} k(\omega, s)F(s)ds,$$

where $F(\omega) = f(\beta(\omega))$, $M(\omega) = m(\beta(\omega))$, $k(\omega, s) = K(\beta(\omega), \beta(s))$. We thus have our main result:

THEOREM 3. *When $G(x, y)$ is absolutely continuous $g(y)$ the Fredholm-Stieltjes integral equation* (1) *is reducible to an ordinary Fredholm integral equation.*

4. **Mixed linear equations.** The mixed equation*

$$f(x) = m(x) + \sum_{i=1}^{m} \lambda K^{(i)}(x)f(s_i) + \lambda \int_0^1 K(x, s)f(s)ds \tag{5}$$

can easily be put into the form

$$f(x) = m(x) + \lambda \int_0^1 R(x, s)f(s)dg(s).$$

Thus from Theorem 3 we see that equation (5) is reducible to a Fredholm integral equation.

DUKE UNIVERSITY

A THEOREM ON QUADRATIC FORMS†

WILLIAM T. REID

In this note the following result is proved:

THEOREM. *Suppose $A[x] \equiv a_{\alpha\beta}x_\alpha x_\beta$,‡ $B[x] \equiv b_{\alpha\beta}x_\alpha x_\beta$ are real quadratic forms in (x_α), $(\alpha = 1, \cdots, n)$, and that $A[x] > 0$ for all real $(x_\alpha) \neq (0_\alpha)$ satisfying $B[x] = 0$. Then there exists a real constant λ_0 such that $A[x] - \lambda_0 B[x]$ is a positive definite quadratic form.*

This theorem is of use in considering the Clebsch condition for multiple integrals in the calculus of variations. A. A. Albert§ has given

* W. A. Hurwitz, *Mixed linear integral equations of the first order*, Transactions of this Society, vol. 16 (1915), pp. 121–133.

† Presented to the Society, December 30, 1937.

‡ The tensor analysis summation convention is used throughout.

§ This Bulletin, vol. 44 (1938), pp. 250–253.

an algebraic proof of this result. The following proof is more analytic in character than that of Albert and is of interest because in the proof itself we obtain directly, in terms of the roots of the characteristic equation (1) below, the interval in which the value λ_0 of the theorem may be chosen.

Suppose that the matrix $\|b_{\alpha\beta}\|$ has rank $n-r$, $(0 \leqq r < n)$, and let $u_\alpha = u_{\alpha\kappa}$, $(\kappa = 1, \cdots, r)$, be linearly independent solutions of the equations $b_{\alpha\beta}u_\beta = 0$, $(\alpha = 1, \cdots, n)$.

LEMMA 1. *The characteristic equation*

$$D(\lambda) \equiv |\, a_{\alpha\beta} - \lambda b_{\alpha\beta} \,| = 0 \tag{1}$$

is a polynomial of degree $n-r$ in λ.

Let $u_{\alpha r+s}$, $(s = 1, \cdots, n-r)$, be sets linearly independent of the above defined $u_{\alpha\kappa}$ and such that the determinant $|u_{\alpha\beta}|$, $(\alpha, \beta = 1, \cdots, n)$, is different from zero. Clearly the degree of $D(\lambda)$ is equal to that of the polynomial $D_1(\lambda) = |(a_{\alpha\gamma} - \lambda b_{\alpha\gamma})u_{\gamma\beta}|$. Since the first r columns of $D_1(\lambda)$ are independent of λ, we obviously have $d^k D_1/d\lambda^k \equiv 0$ if $k > n-r$. Also, the $(n-r)$th derivative is equal to $(n-r)!\,|a_{\alpha\gamma}u_{\gamma\kappa} - b_{\alpha\gamma}u_{\gamma r+s}|$, $(\kappa = 1, \cdots, r;\ s = 1, \cdots, n-r)$; and if its value is zero, there are constants $(c_\alpha) \neq (0_\alpha)$ such that $u_\alpha^0 = u_{\alpha\kappa}c_\kappa$, $u_\alpha' = u_{\alpha r+s}c_{r+s}$ satisfy the equations $a_{\alpha\beta}u_\beta^0 = b_{\alpha\beta}u_\beta'$. By use of these equations together with $u_\alpha^0 b_{\alpha\beta} = 0_\beta$ we have $A[u^0] = 0$; hence $(u_\alpha^0) = (0_\alpha)$, by the hypothesis of the theorem, and consequently $(c_\kappa) = (0_\kappa)$. This implies in turn that $b_{\alpha\beta}u_\beta' = 0_\alpha$; and since $(u_\alpha') \neq (0_\alpha)$ and is independent of the sets $u_{\alpha\kappa}$, $(\kappa = 1, \cdots, r)$, we have a contradiction to the assumption that $\|b_{\alpha\beta}\|$ is of rank $n-r$. The above lemma is therefore proved.

Corresponding to a root of (1) the number of linearly independent solutions of the associated equations

$$(a_{\alpha\beta} - \lambda b_{\alpha\beta})y_\beta = 0_\alpha, \qquad \alpha, \beta = 1, \cdots, n, \tag{2}$$

is termed the *index* of λ as a root of (1). For sets $x_\alpha' \equiv (x_\alpha')$, $x_\alpha'' \equiv (x_\alpha'')$ we denote by $A[x'; x'']$, $B[x'; x'']$ the bilinear forms $a_{\alpha\beta}x_\alpha' x_\beta''$, $b_{\alpha\beta}x_\alpha' x_\beta''$. The first of the following lemmas is immediate.

LEMMA 2. *If λ', λ'' are distinct zeros of $D(\lambda)$, and y_α', y_α'' are solutions of the corresponding equations* (2), *then $B[y'; y''] = 0 = A[y'; y'']$.*

LEMMA 3. *The zeros of $D(\lambda)$ are all real, and the corresponding solutions of* (2) *may be chosen real.*

For suppose $\lambda' = \lambda_1 + i\lambda_2$, $(\lambda_2 \neq 0)$, is a zero of $D(\lambda)$, and $y_\alpha = y_{\alpha 1} + iy_{\alpha 2}$ is a corresponding solution of (2). Then the conjugate values $\bar{y}_\alpha$ satisfy (2) for $\lambda = \bar{\lambda}'$. By Lemma 2 we have $0 = A[y; \bar{y}] = A[y_1] + A[y_2]$, $0 = B[y; \bar{y}] = B[y_1] + B[y_2]$. Now if either $B[y_1]$ or $B[y_2]$ is zero, so also is the other, and, in view of the hypothesis of the theorem, we have a contradiction. On the other hand, if $B[y_2] \neq 0$, let c be a value opposite in sign to $\lambda_2/B[y_1]$ and such that $0 = B[y_1 + cy_2] = B[y_1] + 2cB[y_1; y_2] + c^2 B[y_2] = (1 - c^2)B[y_1] + 2cB[y_1; y_2]$. Using the relations satisfied by $y_{\alpha 1}$, $y_{\alpha 2}$ one obtains, by direct calculation, $A[y_1 + cy_2] = (2c\lambda_2/B[y_1])(\{B[y_1]\}^2 + \{B[y_1; y_2]\}^2) < 0$, which is a contradiction. Hence there are no complex zeros of $D(\lambda)$, and the solutions of equations (2) may be chosen real.

LEMMA 4. *If $\lambda = \lambda'$ is a zero of $D(\lambda)$, then its index is equal to its multiplicity.*

Suppose the index of λ' is equal to k, and denote by $y_\alpha = y_{\alpha j}$, $(j = 1, \cdots, k)$, corresponding linearly independent solutions of (2). Now choose $y_{\alpha h}$, $(h = k+1, \cdots, n)$, such that the determinant $|y_{\alpha\beta}|$, $(\alpha, \beta = 1, \cdots, n)$, is different from zero. Clearly the multiplicity of λ' as a zero of $D(\lambda)$ is equal to its multiplicity as zero of

$$\Delta(\lambda) \equiv |(a_{\alpha\gamma} - \lambda b_{\alpha\gamma}) y_{\gamma\beta}|.$$

Since the first k columns of $\Delta(\lambda)$ are zero for $\lambda = \lambda'$, the multiplicity of λ' as a zero of $\Delta(\lambda)$ is not less than k. Moreover,

$$d^k\Delta/d\lambda^k|_{\lambda=\lambda'} = k!\,|-b_{\alpha\beta}y_{\beta j}\,(a_{\alpha\beta} - \lambda' b_{\alpha\beta})y_{\beta h}|, \qquad j = 1, \cdots, k;\ h = k+1, \cdots, n. \tag{3}$$

Now if this derivative is zero, there are constants $(c_\alpha) \neq (0_\alpha)$ such that the set $y_\alpha^0 = y_{\alpha j} c_j$, $y_\alpha' = y_{\alpha h} c_h$ satisfies

$$(a_{\alpha\beta} - \lambda' b_{\alpha\beta}) y_\beta' = b_{\alpha\beta} y_\beta^0. \tag{4}$$

Since (y_α^0) satisfies (2) for $\lambda = \lambda'$ we find from (4) that $B[y^0] = 0$, and consequently, using again equations (2), that $A[y^0] = 0$. In view of the hypothesis of the theorem it then follows that $(y_\alpha^0) = (0_\alpha)$ and $(c_j) = (0_j)$. But from equations (4) we see that (y_α') is then a solution of (2) for $\lambda = \lambda'$; and, since this set is independent of the k sets $y_{\alpha j}$, we have a contradiction to the assumption that k is the index of λ'. Hence the expression (3) is not zero, and the multiplicity of λ' as a zero of $D(\lambda)$ is also k.

Now let $\lambda_1 \leqq \lambda_2 \leqq \cdots \leqq \lambda_{n-r}$ denote the zeros of $D(\lambda)$ each repeated a number of times equal to its index, and let $y_\alpha = y_{\alpha s}$, $(s = 1, \cdots, n-r)$, be corresponding linearly independent real solutions of equations (2).

LEMMA 5. *The solutions $y_{\alpha s}$ are such that $B[y_s] \neq 0$, $(s = 1, \cdots, n-r)$; moreover, these solutions may be so chosen that $B[y_s; y_t] = 0$ if $s \neq t$.*

The first part of the lemma follows readily from the hypothesis of the theorem, since $A[y_s] = \lambda_s B[y_s]$, $(s = 1, \cdots, n-r)$. If $\lambda_s \neq \lambda_t$, the second part of the lemma is true by Lemma 2. If $\lambda_s = \lambda_t$, the corresponding solutions may be so orthogonalized in view of the relation $B[y_s] \neq 0$.

COROLLARY. *The determinant $|u_{\alpha s} y_{\alpha s}|$ is different from zero.*

LEMMA 6. *If for a value s we have $B[y_s] > 0$, then for each t such that $\lambda_t \geqq \lambda_s$ we also have $B[y_t] > 0$.*

For suppose $\lambda_t \geqq \lambda_s$ and $B[y_t] < 0$. If $c^2 = -B[y_t]/B[y_s]$, then by the above lemma we have $(y_{\alpha t} + c y_{\alpha s}) \neq (0_\alpha)$. Moreover, $B[y_t + c y_s] = B[y_t] + c^2 B[y_s] = 0$, $A[y_t + c y_s] = A[y_t] + c^2 A[y_s] = B[y_t](\lambda_t - \lambda_s) \leqq 0$, which is impossible in view of the hypothesis of the theorem.

Now let Λ denote the interval of the λ axis defined by $\lambda < \lambda_1$ if $B[y_1] > 0$, $\lambda > \lambda_{n-r}$ if $B[y_{n-r}] < 0$, $\lambda_g < \lambda < \lambda_{g+1}$ if $B[y_g] < 0$, $B[y_{g+1}] > 0$. The interval Λ is uniquely determined in view of Lemma 6. We shall now show that $A[x] - \lambda_0 B[x]$ is a positive definite quadratic form for λ_0 on Λ. If $(x_\alpha) \neq (0_\alpha)$, it follows from the corollary to Lemma 5 that there are unique values $(c_s, d_s) \neq (0_s, 0_s)$ such that $x_\alpha = u_{\alpha s} c_s + y_{\alpha s} d_s$. Moreover, since $u_{\alpha s} b_{\alpha\beta} = 0$, we have $A[u_s; y_s] = 0$, $(s = 1, \cdots, r;\ s = 1, \cdots, n-r)$, and $A[u_s, u_r] c_s c_r \geqq 0$, with the inequality sign holding unless $(c_s) = (0_s)$. Now if λ_0 is on the above defined interval Λ, each of the terms $(\lambda_s - \lambda_0) B[y_s]$ is positive. From the relations

$$A[x] - \lambda_0 B[x] = A[u_s; u_r] c_s c_r + (A[y_s; y_t] - \lambda_0 B[y_s; y_t]) d_s d_t$$

$$= A[u_s; u_r] c_s c_r + \sum_{s=1}^{n-r} (\lambda_s - \lambda_0) B[y_s] d_s^2$$

we therefore conclude that $A[x] - \lambda_0 B[x]$ is positive definite for λ on Λ. In conclusion it is to be noted that if λ_0 is not on this interval, the form $A[x] - \lambda_0 B[x]$ is negative or zero for at least one of the sets $(x_\alpha) = (y_{\alpha s})$, $(s = 1, \cdots, n-r)$.

UNIVERSITY OF CHICAGO

ON THE EXTENDED FORM OF CAUCHY'S CONDENSATION TEST FOR THE CONVERGENCE OF INFINITE SERIES

R. E. MORITZ

Cauchy's celebrated theorem,* which forms the basis of some of the most powerful chain tests for the convergence and divergence of infinite series, notably the logarithmic scale and DeMorgan and Bertrand's T and τ scales, is as follows:

Let $f(n)$ represent any positive continuous and single-valued function of n which is monotone decreasing as n increases. Suppose, moreover, that $\lim_{n\to\infty} f(n)=0$. Then $\sum^{\infty} f(n)$ and $\sum^{\infty} a^n f(a^n)$ are both convergent or both divergent, provided a is a positive integer not less than 2.

The restriction $a \leqslant 2$ was removed by Kohn† in 1882. The proof covers several pages. Kohn showed that the theorem holds for any real value of a greater than unity.

A shorter proof of Kohn's extension of Cauchy's theorem was given by Hill‡ in 1896. Hill's proof depends on two auxiliary theorems which, in turn, are based on the two previously established inequalities

$$\left(1+\frac{1}{a}+\frac{1}{a^{s+1}}\right) t_{s+1} < t_{s+1} - t_s,$$

$$t_{s+1} - t_s < \left(a - 1 + \frac{1}{a^s}\right)(1 + t_s),$$

where t_s is the integral part of a^s.

In view of the fundamental importance of Cauchy's theorem the following simple analytical proof of the theorem with Kohn's extension may, therefore, be of some interest. The proof which follows is in fact in the nature of a corollary to the well known integral test which may be stated as follows:

Under the assumptions set forth in Cauchy's theorem, $\sum^{\infty} f(n)$ and $\int^{\infty} f(x)dx$ converge and diverge together.

Set $x=a^u$, $(a>1)$; then $\log a>0$ and

---

* Cauchy, Anciens Exercises, vol. 2 (1827), p. 221.

† Kohn, G., Grunert Archiv, vol. 67 (1882), pp. 63–95.

‡ Hill, M. J. M., Messenger of Mathematics, (2), vol. 26 (1896–1897), p. 102.

$$\int^{\infty} f(x)dx = \log a \int^{\infty} a^u f(a^u)du = \log a \int^{\infty} a^x f(a^x)dx.$$

Therefore $\sum^{\infty} f(n)$ converges and diverges with $\int^{\infty} a^x f(a^x)dx$, and it remains only to show that $\int^{\infty} a^x f(a^x)dx$ converges and diverges with $\sum^{\infty} a^n f(a^n)$.

For values of x between the two integers n and $n+1$, we have, if a is greater than unity, $a^n < a^x < a^{n+1}$; and since $f(x)$ is monotone decreasing,

$$f(a^{n+1}) < f(a^x) < f(a^n).$$

Furthermore, since $f(x)$ is positive we have, on multiplying these inequalities term by term,

$$a^n f(a^{n+1}) < a^x f(a^x) < a^{n+1} f(a^n),$$

from which, on integrating with respect to x between the limits n and $n+1$, we have

$$\frac{1}{a} \cdot a^{n+1} f(a^{n+1}) < \int_n^{n+1} a^x f(a^x)dx < a \cdot a^n f(a^n)$$

and therefore

$$\frac{1}{a} \sum_{n+1}^{\infty} a^k f(a^k) < \int_n^{n+1} a^x f(a^x)dx + \int_{n+1}^{n+2} a^x f(a^x)dx + \int_{n+2}^{n+3} a^x f(a^x)dx + \cdots = \int_n^{\infty} a^x f(a^x)dx < a \sum_n^{\infty} a^k f(a^k).$$

These two inequalities make it obvious that $\sum^{\infty} a^n f(a^n)$ and $\int^{\infty} a^x f(a^x)dx$ converge and diverge together.

This completes the proof of both Cauchy's theorem and Kohn's extension.

UNIVERSITY OF WASHINGTON

ON THE BOUNDARY CONDITION $\partial u/\partial n + au = 0$ FOR HARMONIC FUNCTIONS*

HILLEL PORITSKY

1. **Introduction.** In a recent paper (referred to below as I)† the author considered the nature of the "reflection" or analytic continuation of harmonic functions across a plane over which the boundary condition

$$\frac{\partial u}{\partial n} + au = 0, \qquad a = \text{const.}, \tag{1}$$

applies. In the following this investigation is continued. First, the case of a spherical (circular in two dimensions) boundary, over which the boundary condition (1) holds is considered. It is shown that singularities admit of a similar "reflection," as in the case of a plane in I; thus a point singularity at P_0 outside a sphere S over which (1) holds is reflected into a point singularity at P_1 (the spherical inverse of P_0 in S) and into a distribution of singularities along the radius vector from the center 0 to P_1.

Returning to plane boundaries, we apply boundary conditions of the form (1) over *two parallel* planes. A rather complicated "reflection" of singularities results, consisting of point singularities as well as of distributed line singularities. The point singularities are located at the periodic row of points obtained by reflecting the original singularity P_0 first in one plane, then in the other one; reflecting these images in the two planes; and so forth. The line singularities are distributed over the straight line through the above point singularities. The density of the distributed line singularities is an analytic function of the distance along the line bearing them between the point singularities, but changes abruptly from one analytic function to another one in crossing these points.

Some of the above features are believed to be typical of analytic continuations of a great variety of expansions in characteristic functions related to two point problems.

2. **Circular and spherical boundaries.** We shall consider the question of reflections of singularities of harmonic functions across spherical or circular boundaries corresponding to the condition (1).

* Presented to the Society, September 5, 1934.

† This Bulletin, vol. 43 (1937), pp. 873–885.

First we consider the two-dimensional harmonic case. Suppose that the function u satisfies

$$\frac{\partial u}{\partial r} - au = 0, \quad \text{for} \quad r = 1, \tag{2}$$

and is harmonic outside of this circle except near a point P_0 where it becomes infinite like $-\ln r_0$; u is also to vanish at infinity. The point P_0 is along the x axis at $x = h = e^H > 1$. In terms of (two-dimensional) potential theory language, the singularity of u at P_0 is that of a "unit point charge."

The determination of u and its analytic continuation may be carried out by reducing it to the case considered in I by the application of the conformal transformation

$$z = e^Z, \qquad Z = X + iY, \qquad z = x + iy. \tag{3}$$

This straightens out the boundary $r = 1$ by transforming it into the line $X = 0$; the boundary condition (2) is preserved, while the unit point charge at $z = e^H$ is carried into unit point charges at

$$Z = \ln h = H + 2n\pi i. \tag{4}$$

Proceeding as in I, §5, we find that each one of the periodic array of point charges is "reflected" into a point charge and an exponential trail of negative charges. When we change back to the z plane, there results a positive image at $z = 1/h$ and a distributed charge density

$$\rho(x) = -2ah^a x^{a-1}, \quad \text{for} \quad 0 < x < h. \tag{5}$$

In determining the latter, it is to be kept in mind that charge is preserved under conformal transformations, so that $\rho(x)$ is found from $\rho(x)dx = P(X)dX$, where $P(X)$ is the charge density along the X axis in the Z plane. From I, equation (23), $P(X) = -2ae^{a(h+z)}$, for $x < -h$.

As in I, the restriction I, (9), is pertinent. The total amount of charge then vanishes so that u vanishes at infinity like $O(|z|^{-1})$.

Proceeding to the three-dimensional case and a spherical boundary along which (2) holds, we can no longer utilize conformal transformations but proceed as follows.

In order to determine the function u which is harmonic except near $P_0: (x, y, z) = (h, 0, 0)$, $(h > 1)$, where u becomes infinite like $1/r_0$ (the reciprocal distance from P_0) while along the unit sphere $r = 1$ it satisfies the boundary condition (2), start with the expansion

$$u = \sum_{n=0}^{\infty} A_n r^{-(n+1)} P_n(\cos\theta) + 1/r_0, \tag{6}$$

for u, $(r \geqq 1)$, and utilize the familiar expansions

$$\frac{1}{r_0} = \left[\sum_{n=0}^{\infty} \frac{r^n}{h^{n+1}} P_n(\cos\theta) \quad \text{or} \quad \sum_{n=0}^{\infty} \frac{h^n}{r^{n+1}} P_n(\cos\theta)\right], \tag{7}$$

according as r is less than or greater than h, where θ is the angle with the x axis and P_n are the Legendre polynomials. If (2) is imposed,

$$A_n = \frac{1}{h^{n+1}} - \frac{(1+2a)}{(n+a+1)h^{n+1}}. \tag{8}$$

The first part of A_n leads to a point charge of amount $1/h$ at P_1, the spherical inverse point of P_0; the latter leads to a distributed charge density $\rho(x)$ along the line from the center to P_1, provided $\rho(x)$ satisfies the moment equations

$$\int_0^{1/h} \rho(x')(x')^n dx' = -\frac{(1+2a)}{(n+a+1)h^{n+1}}, \qquad n = 0, 1, 2, \cdots. \tag{9}$$

The solution of these equations is given by

$$\rho(x) = \frac{(1+2a)x^a}{h^a}. \tag{10}$$

Magnification can be used for treating the case where (2) holds along a sphere of non-unit radius, while singularities inside the circle or sphere can be reduced to outer singularities by means of inversion.

The two-dimensional case can also be treated in a similar fashion with Fourier series replacing the series of Legendre polynomials.

3. **Boundary conditions along two parallel planes.** We consider in this section continuations of harmonic functions across two parallel planes $x=0$, $x=C$, corresponding to boundary conditions of the form (1) on each plane, though for simplicity we suppose that (1) applies along $x=0$. Then

$$\frac{\partial u}{\partial x} - au = 0, \qquad \text{for} \quad x = 0, \tag{11}$$

$$\frac{\partial u}{\partial x} = 0, \qquad \text{for} \quad x = C. \tag{12}$$

We shall investigate the reflections of a point charge singularity located at P_0: $(h, 0, 0)$ between the planes $x=0$, $x=C$, $(0<h<C)$.

One way of obtaining the singularities of the analytic continuation is to apply the familiar method of successive reflections, attending

alternately first to $x=0$, then to $x=C$. There result point charges along the x axis at

$$x = \pm h \pm 2nC, \qquad n = 0, \pm 1, \pm 2, \cdots, \tag{13}$$

as well as proper continuous charge distributions which change abruptly to different analytic functions on crossing the points (13). This process of reflection gets rather irksome after several stages.

We shall first follow these reflections by means of the operational treatment employed in I, §8. If $S_0(x-h)$ is the original charge at $x=h$, reflection across $x=C$ leads to the new charge $S_0(x-2C+h)$. As shown in I, §8, analytic continuation across $x=0$ results in multiplication by the operator $(p+a)/(p-a)$, in addition to replacement of the argument by its negative. This introduces the charges

$$\frac{p+a}{p-a}\left[S_0(x+h) + S_0(x+2C-h)\right]. \tag{14}$$

Further reflection in $x=C$ replaces x by $2C-x$, changes $f(p)$ to $-f(-p)$, and results in the charges

$$\frac{a-p}{a+p}\left[S_0(x-2C-h) + S_0(x-4C+h)\right]. \tag{15}$$

Following this by a reflection in $x=0$ we would apparently have

$$\frac{p+a}{p-a}\frac{a-p}{a+p}\left[S_0(x+2C+h) + S_0(x+4C-h)\right] = -\left[S_0(x+2C+h) + S_0(x+4C-h)\right]. \tag{16}$$

This, however, may be shown to be incorrect. The explanation of the difficulties thus encountered by the operational treatment will appear presently.

We shall give a Bessel-Fourier integral representation for u. Utilizing I, equation (16), we put

$$u = \int_0^\infty J_0(\lambda\rho)\left[e^{-\lambda|x-h|} + f(\lambda)e^{\lambda x} + g(\lambda)e^{-\lambda x}\right]d\lambda, \tag{17}$$

where $f(\lambda)$, $g(\lambda)$ are to be determined. Imposing the conditions (11), (12) on the integrand and solving, say for $f(x)$, we obtain

$$f(\lambda) = \frac{e^{+\lambda h}\dfrac{\lambda+u}{\lambda-a} + e^{-\lambda h}}{e^{2\lambda C}\dfrac{\lambda+a}{\lambda-a} - 1}. \tag{18}$$

Expanding $f(\lambda)$ in powers of $e^{-2\lambda C}(\lambda-a)/(\lambda+a)$ one obtains

$$(19)\quad f(\lambda) = e^{\lambda(h-2C)}\sum_{n=0}^{\infty}\left(\frac{\lambda-a}{\lambda+a}\right)^n e^{-2\lambda nC} + e^{-\lambda h}\sum_{n=0}^{\infty}\left(\frac{\lambda-a}{\lambda+a}\right)^{n+1} e^{-2\lambda nC},$$

and the substitution of (19) in the middle term of (17) results in the potential function

$$(20)\quad \sum_{n=0}^{\infty}\int_0^{\infty}\left(\frac{\lambda-a}{\lambda+a}\right)^n e^{\lambda[x+h-2(n+1)C]}J_0(\lambda\rho)d\lambda + \sum_{n=0}^{\infty}\int_0^{\infty}\left(\frac{\lambda-a}{\lambda+a}\right)^{n+1} e^{\lambda(x-h-2nC)}J_0(\lambda\rho)d\lambda.$$

To interpret (20) as the potential of charges, note that from I, equation (16), it follows that a charge distribution $c(x)$ over the positive x axis gives rise to the potential

$$(21)\quad \int_0^{\infty} f(\lambda)e^{\lambda x}J_0(\lambda\rho)d\lambda, \qquad x<0,$$

where

$$(22)\quad f(\lambda) = \int_0^{\infty} c(x)e^{-\lambda x}dx.$$

Thus, if $f(\lambda)$ is given (say by means of (18) or (19)), one obtains for $c(x)$ a Laplace-Carson integral equation. The solution of the latter can be effected by means of a Bromwich integral; in operational notation the solution appears in the form

$$(23)\quad c(x) = f(p)S_0(x).$$

The explicit charge corresponding to the displayed term in the first series in (20) is thus

$$(24)\quad \left(\frac{p-a}{p+a}\right)^n e^{p[h-2(n+1)C]}S_0(x).$$

The effect of the exponential is to translate the distribution a distance $2(n+1)C-h$ in the direction of positive x; there remains to interpret

$$(25)\quad \left(\frac{p-a}{p+a}\right)^n S_0(x) = \left(1-\frac{2a}{p+a}\right)^n S_0(x).$$

This is done by expanding by the binomial theorem and replacing $(p+a)^{-k}S_0(x)$ by $x^{k-1}e^{-ax}/(k-1)!$, for $x>0$, and by 0 for $x<0$. Of

course, the first term in the expansion of the right-hand side of (25) leads to a unit point charge at $x=0$. It follows that (24) represents a point charge at $x=2(n+1)C-h$ and a continuous charge beyond this point of charge density $P_n(x)e^{-ax}$, where $P_n(x)$ is a polynomial of degree n.

As regards the function $g(\lambda)$ and the last integral in (17), a similar treatment is possible resulting in an interpretation in terms of charges along the negative x axis. The latter might be described essentially as the positive image of the charges for $x>C$ in the plane $x=C$. However, in place of (22) an integral equation of the form

$$g(\lambda) = \int_{-\infty}^{0} c_1(x)e^{\lambda x}dx \tag{26}$$

now obtains for the charge density $c_1(x)$ along $x<0$. A Bromwich integral will not yield the solution of this equation unless one changes the independent variable to $-x$ or properly changes the path of integration.

An operational solution of (26) (similar to the solution (23) of (22)) could therefore be given only by using a non-orthodox interpretation of the operator $g(p)$. Essentially the same situation appeared in I, where, in equation (42), $1/(p-\alpha)S_0(x)$ was interpreted as $-e^{\alpha x}H_0(-x)$ rather than $e^{\alpha x}H_0(x)$, as is usually done in operational calculus* (see also I, footnote, p. 884). This also accounts for the difficulties encountered in connection with (16) and the attempted operational treatment of the successive reflections (two distinct (and mutually inconsistent) interpretations of the operators $1/(p-a)$, $1/(p+a)$ occurring in (15), (16) have to be employed).†

A very similar treatment can be given for the case wherein general linear conditions of the form (11) apply along both $x=0$ and $x=C$.

We believe that some of the features of the above analytic continuation are characteristic of general differential systems with two point boundary conditions. It is planned to touch upon this phase in the future.

General Electric Co.
Schenectady, N. Y.

* Both interpretations yield the Green's function of the differential equation $(d/dx-\alpha)u=0$ with discontinuity at $x=0$, the one relative to the boundary condition $u(+\infty)=0$, the other relative to $u(-\infty)=0$.

† Roughly speaking, the operational method fails here because of its catering to functions which vanish for sufficiently negative x. It can handle the charges for $x>0$. Properly modified (as in I, §8) it can handle the charges for $x<0$. It is not adapted, however, to treating both sets of charges and exhibiting their relations.

ON GENERATING THE SIMPLE GROUP $LF(2, 2^N)$ BY TWO OPERATORS OF PERIODS TWO AND THREE*

ABRAHAM SINKOV

The purpose of this paper is to consider the question of the number of abstractly distinct ways in which it is possible for two operators of periods two and three to generate the simple group $LF(2, 2^N)$. The general procedure to be followed in studying such a problem has been outlined by Professor Brahana† and has already been applied to $LF(2, 2^3)$ by the author.‡ Those previous results suggested the present generalization.

Since the $2^{2N}-1$ substitutions of period two in $LF(2, 2^N)$ are all conjugate,§ it is sufficient in seeking possible generating operators to consider only one of them, say

$$T \equiv \left(\frac{1, 1}{0, 1}\right).$$

In the representation of G on 2^N+1 letters, T leaves fixed the single element ∞. The largest subgroup within which T is invariant is $G^{(\infty)}$, composed of all the substitutions

$$T_\mu \equiv \left(\frac{1, \mu}{0, 1}\right), \qquad \mu \text{ in the } GF[2^N],$$

which keep the single element ∞ unchanged. The group $G^{(\infty)}$ is abelian, of order 2^N and of type $(1, 1, 1, \cdots)$.

If the operators of period three are divided up into sets of complete conjugates under $G^{(\infty)}$, then the various members of each set satisfy with T the same abstract relations, and it is sufficient to select from each such set only one operator to serve as a possible second generator with T. The number of these sets to be considered depends

* Presented to the Society, February 20, 1937.

† H. R. Brahana, *Pairs of generators for the known simple groups whose orders are less than one million*, Annals of Mathematics, vol. 31 (1930), pp. 542, 543.

‡ *Necessary and sufficient conditions for generating certain simple groups by two operators of periods two and three*, American Journal of Mathematics, vol. 49 (1937), pp. 69–72, hereafter referred to as *Necessary and sufficient conditions*.

§ The various properties of $LF(2, 2^N)$ of which use is made in this paper are given in Dickson's *Linear Groups with an Exposition of the Galois Field Theory*, 1901, chap. 12.

on the form of 2^N+1. For, when 2^N+1 is a multiple of three, G contains $2^N(2^N-1)$ operators of period three, whereas when 2^N+1 is not a multiple of three, there are $2^N(2^N+1)$ such operators.

In the former case, there are 2^N-1 sets of 2^N operators each, and we may choose as a representative from each set the operator

$$S_\alpha \equiv \begin{pmatrix} 0, & x^{2\alpha} \\ 1, & x^\alpha \end{pmatrix}, \qquad \alpha = 0, 1, 2, \cdots, 2^N - 2,$$

where x is a primitive root of the $GF[2^N]$. In the latter case, there are 2^N+1 sets of 2^N operators each. We may again choose the 2^N-1 operators S_α together with the two additional operators

$$S_{2^N} \equiv \begin{pmatrix} x^{(2^N-1)/3}, & 0 \\ 0, & 1 \end{pmatrix}$$

and $S_{2^N+1}=(S_{2^N})^2$. Since

$$S_{2^N}T = \begin{pmatrix} x^{(2^N-1)/3}, & x^{(2^N-1)/3} \\ 0, & 1 \end{pmatrix}$$

is of period three, S_{2^N} and T generate a tetrahedral group of order 12. The same is true of S_{2^N+1} and T, so that neither of these pairs can be used to generate the entire group. Moreover, S_0T is of period two, so that $\{S_0, T\}$ is of order six. In every case then, it is sufficient to consider only the 2^N-2 operators S_α, $(\alpha \neq 0)$.

Now it may happen, for a particular subscript k, that T and S_k generate a proper subgroup H of G instead of the entire group. We note however that

$$S_\alpha T = \begin{pmatrix} 0, & x^{2\alpha} \\ 1, & 1 + x^\alpha \end{pmatrix}$$

may not be of period 2, 3, or 4 for any α whatever. Hence, on referring to the list of possible subgroups of $LF(2, 2^N)$, one sees that H is either a linear fractional group $LF(2, 2^{N/r})$ or else is of order $2^s d$. But the latter case may be ruled out. For, if a subgroup H of order $2^s d$ existed, the commutative subgroup within it of order 2^s would include the commutator subgroup of H. Hence d, and therefore the period of $S_\alpha T$, is at most three,* which is not possible.

H is therefore a linear fractional group, and the number x^k must be an element in the $GF[2^{N/r}]$. If we set $N/r=t$, then $(x^k)^{2^t-1}=1$, from which $k(2^t-1)\equiv 0$ modulo (2^N-1). The number k is thus re-

* H. R. Brahana, *Certain perfect groups generated by two operators of periods two and three*, American Journal of Mathematics, vol. 50 (1928), p. 348.

quired to be a multiple of $(2^N-1)/(2^t-1)$. This condition is also sufficient, since every subscript which has, with 2^N-1, a greatest common divisor of the form $(2^N-1)/(2^t-1)$, t being a divisor of N, corresponds to an operator S which generates with T a subgroup of $LF(2, 2^t)$ and hence a proper subgroup of $LF(2, 2^N)$.

We may therefore eliminate from the set of operators S_α all those whose subscripts satisfy the above requirement. There will then remain, in place of the original 2^N-2, exactly

$$A \equiv 2^N - \sum 2^{N/p_1} + \sum 2^{N/p_1p_2} - \cdots + (-1)^r \sum 2^{N/p_1p_2\cdots p_r},$$

where the p_i are the distinct prime divisors of N and the jth summation is taken with respect to all the possible combinations of these p_i, j at a time. This number A is a simplification of the expression

$$(2^N - 2) - \sum (2^{N/p_1} - 2) + \sum (2^{N/p_1p_2} - 2) - \cdots + (-1)^r \sum (2^{N/p_1p_2\cdots p_r} - 2),$$

and its correctness is established as follows: Let k be of the required form. Then if t involves b distinct primes, S_k will appear $C_{b,j}$ times in the jth summation. The number of times that it will appear in the entire expression is

$$1 - b + C_{b,2} - \cdots + (-1)^b = (1-1)^b \equiv 0.$$

Any subscript not of the form required by the preceding paragraph will appear only once in the first term.

The operators S which now remain will all generate with T the entire group G. But the pairs are not necessarily abstractly distinct, for in reducing their number to A we have taken account of only the inner automorphisms of G. It still remains necessary to consider the possible outer automorphisms. The group of automorphisms of the linear fractional group has been studied by O. Schreier and B. L. van der Waerden.* It follows from their results that the only possible outer automorphisms of $LF(2, 2^N)$ are to be found among the automorphisms of the $GF[2^N]$. From this result, it can be seen that the group of outer automorphisms of $LF(2, 2^N)$ is the cyclic group generated by the substitution $z' = z^2$, of period N.

This substitution is commutative with T and transforms S_l into S_{2l}, provided we consider the subscript $2l$ reduced modulo (2^N-1). As a result, the relations satisfied by T and S_l are abstractly identical with those satisfied by T and S_{2^il}, where i may take on any value

* Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität, vol. 6 (1928), pp. 303–322.

whatever. Hence the A operators S may be divided up into sets, in such a way that the subscripts of all the operators in any set are obtainable from some one of them by continued multiplications by two and reduction modulo (2^N-1). The number of operators in any particular set is exactly N, except in those cases when the number 2^N-1 and the subscripts of the set have a greatest common divisor of the form $(2^N-1)/(2^r-1)$. But these have already been eliminated. The number of sets obtained, and therefore the number of distinct definitions of the group, is consequently A/N.

As a corollary, we have the number theoretic result that A/N is always an integer. When N is prime, the corollary reduces to a special case of Fermat's theorem. It was pointed out to me by Professor Gill that the number A/N is identically Dickson's $N_{N,2}$,* or the number of irreducible polynomials of degree N in a $GF[2]$. This coincidence can be used to advantage in the following way:

Consider two distinct irreducible polynomials in the $GF[2]$ of degree N in the variables x and X, respectively. Let S_1 and σ_1 designate the substitutions

$$\begin{pmatrix} 0, & x^2 \\ 1, & x \end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix} 0, & X^2 \\ 1, & X \end{pmatrix},$$

respectively, and suppose that the pairs of generators T, S_1 and T, σ_1 are abstractly identical. Then we get an automorphism of G by making T correspond to itself and S_1 correspond to σ_1. As a result, every combination of S_1 and T will have the same period as the corresponding combination of σ_1 and T. Consider in particular all the combinations of the generators which will reduce to the identity. The formal procedure of calculating them is identical for the two separate cases if we hold all processes of modular reduction by means of the polynomial modulus until the very last step. We see then that we can thus obtain a considerable number of pairs of identical expressions in x and X such that both members of each pair will reduce to the same value. But such a result will necessarily imply that the two polynomial moduli are identical, which is a contradiction. Hence a given S_a will yield with T all the possible definitions of $LF(2, 2^N)$ if we take advantage of all the distinct methods of generating the $GF[2^N]$.

We shall bring this paper to a close by demonstrating an interesting general property of every pair of generators S and T of an $LF(2, 2^N)$.

In the study of groups generated by two operators of periods two

* *Linear Groups*, p. 18.

and three respectively, the following substitution has been found very useful:

$$P = (ST)^{-1}, \qquad S = P^2Q,$$
$$Q = (ST)^2S, \qquad T = P^2Q.$$

It replaces the relations $S^2 = T^2 = (ST)^n = (S^{-1}T^{-1}ST)^p = 1$ by $P^n = Q^p = (QP^2)^2 = (QP^2)^3 = 1$. We shall designate either of these sets of relations by the abbreviated notation $(2, 3, n; p)$.

From the substitution of the equivalents for S and T in the definitions of P and Q, it results that

$$P = \left(\frac{x^\alpha + x^{2\alpha},\ x^{3\alpha}}{x^\alpha,\ 0}\right) = \left(\frac{1 + x^\alpha,\ x^{2\alpha}}{1,\ 0}\right),$$
$$Q = \left(\frac{x^{2\alpha} + x^{3\alpha},\ x^{3\alpha}}{1,\ x^\alpha + x^{2\alpha} + x^{3\alpha}}\right).$$

By induction

$$Q^{2^N} = \left(\frac{A_N,\ x^{3\alpha}}{1,\ x^\alpha + A_N}\right),$$

where $A_N = \sum_{i=0}^{N+1} x^{(2i+1)\alpha}$. Now the period of Q is the same as that of the commutator of S and T. Since

$$K \equiv TST^{-1}S^{-1} = \left(\frac{1 + x^{2\alpha},\ x^{2\alpha}}{1,\ x^{2\alpha}}\right)$$

and

$$K^{2^N} = \left(\frac{B_N,\ x^{2\alpha}}{1,\ 1 + B_N}\right),$$

where $B_N = 1 + \sum_{i=1}^{N+1} x^{2i\alpha}$, it follows that

$$K^{2^N+1} = \left(\frac{B_N + x^{2\alpha}B_N + x^{2\alpha},\ x^{2\alpha}B_N + x^{4\alpha}}{x^{2\alpha} + B_N,\ x^{2\alpha}B_N}\right)$$

and

$$K^{2^N-1} = \left(\frac{x^{2\alpha}(1 + B_N),\ x^{2\alpha}(1 + x^{2\alpha} + B_N)}{1 + x^{2\alpha} + B_N,\ 1 + B_N(1 + x^{2\alpha})}\right).$$

If the period p of K divides $2^N + 1$, $B_N = x^{2\alpha}$; if it divides $2^N - 1$, $B_N = 1 + x^{2\alpha}$.

Suppose that p divides 2^N+1. Then, since $B_{N-1}^2=B_N+x^{2\alpha}$, we have $B_{N-1}=0$, and

$$A_{N-1} = x^\alpha(B_{N-1}+1) + x^{2\alpha} = x^\alpha + x^{2\alpha},$$

$$Q^{2^{N-1}} = \left(\frac{x^\alpha + x^{2\alpha},\ x^{3\alpha}}{1,\ x^{2\alpha}}\right),$$

$$Q^{2^{N-1}+1} = \left(\frac{x^{5\alpha},\ x^{6\alpha}}{x^{3\alpha},\ x^{4\alpha}+x^{5\alpha}}\right)\quad\left(\frac{x^{2\alpha},\ x^{3\alpha}}{1,\ x^\alpha + x^{2\alpha}}\right),$$

$$Q^{2^{N-1}+1}P = \left(\frac{x^{3\alpha},\ x^{5\alpha}}{x^\alpha + x^{3\alpha},\ x^{3\alpha}}\right) = \left(\frac{x^{2\alpha},\ x^{4\alpha}}{1+x^{2\alpha},\ x^{2\alpha}}\right),$$

$$(Q^{2^{N-1}+1}P)^2 = 1,$$

$$(Q^{(p+1)/2}P)^2 = 1.$$

In the same way, if p divides 2^N-1,

$$Q^{2^{N-1}} = \left(\frac{x^{2\alpha},\ x^{3\alpha}}{1,\ x^\alpha + x^{2\alpha}}\right),$$

and

$$Q^{2^{N-1}}P = \left(\frac{x^{2\alpha},\ x^{4\alpha}}{1+x^{2\alpha},\ x^{2\alpha}}\right)$$

is of period two. Once again, then, $(Q^{(p+1)/2}P)^2=1$. Hence, regardless of what particular value p may have, the relation $(Q^{(p+1)/2}P)^2=1$ is always satisfied. This result has some interesting consequences.

The relations $(2, 3, n; p)$, $(Q^{(p+1)/2}P)^2=1$ define the group designated by H. S. M. Coxeter* as $G^{3,n,p}$. (In this definition, the relation $Q^p=1$ is redundant† and may be omitted.) Because of the symmetry of $G^{3,n,p}$, it follows that if $n\neq p$, there must exist a pair of generators satisfying $(2, 3, p; n)$, $(Q^{(n+1)/2}P)^2=1$. Hence, if we consider all the abstractly distinct definitions of $LF(2, 2^N)$ in terms of two generators of periods two and three, it follows that the totality of values assumed by n is identical with the totality of values assumed by p.

Moreover, suppose that by the adjunction of suitable conditions we could get a complete definition of $LF(2, 2^N)$ in terms of $(2, 3, n; p)$, $(Q^{(p+1)/2}P)^2=1$; and suppose further that $n\neq p$. Let us introduce the substitution $P^2=\bar{Q}$, $Q=\bar{P}^2$ with the additional requirement that $\bar{P}^p=1$. Then $P=\bar{Q}^{(n+1)/2}$, $\bar{P}=Q^{(p+1)/2}$, and the relations

* *The abstract groups* $G^{m,n,p}$, to appear in the Transactions of this Society.

† *Necessary and sufficient conditions*, p. 69.

$$P^n = Q^p = (QP^3)^2 = (QP^2)^3 = (Q^{(p+1)/2}P)^2 = (Q^{(p+3)/2}P^2)^2 = 1$$

become

$$\overline{Q}^n = \overline{P}^p = (\overline{Q}^{(n+3)/2}\overline{P}^2)^2 = (\overline{Q}\overline{P}^2)^3 = (\overline{Q}^{(n+1)/2}\overline{P})^2 = (\overline{Q}\overline{P}^3)^2 = 1,$$

in which the roles of P and Q have been interchanged. We are thus enabled to pass directly to a second complete definition in terms of $(2, 3, p; n)$, $(Q^{(n+1)/2}P)^2 = 1$.

It is interesting to observe the first few special cases of $LF(2, 2^N)$ to see how the results of this paper apply.

When $N = 2$, the group is the icosahedral group $G^{2,5,5}$ which is completely defined by the relations $(2, 3, 5)$. This is the only definition; hence p must be equal to n, a known result.

When $N = 3$, the group* is the simple group of order 504. It has two definitions, based on $(2, 3, 7; 9)$ and $(2, 3, 9; 7)$. Since $G_{504} \equiv G^{2,7,9}$ it is sufficient in each case to add the one relation $(Q^{(p+1)/2}P)^2 = 1$ in order to obtain a complete definition of the group.

No complete definition in terms of two generators has yet been obtained for any $N > 3$. However, Todd has given a complete definition† of $LF(2, 2^N)$ in terms of $N+2$ generators. Since U and R in his definition satisfy $(2, 3, 2^N - 1)$, it follows from the foregoing that these two operators suffice to generate the entire group, with the single exception of the case $N = 2$, and that one of the values assumed by n is $2^N - 1$.

When $N = 4$, there are three definitions, and the paired values of n and p are 15, 17; 17, 15; 17, 17.

As a last instance we mention $N = 5$. Here the six paired values of n and p are 31, 31; 31, 31; 33, 33; 33, 33; 31, 11; 11, 31. The duplication of the pair 31, 31 means that $LF(2, 2^5)$ has two distinct abstract definitions based on $(2, 3, 31; 31)$. Or, to express it differently, $LF(2, 2^5)$ is obtainable in two abstractly distinct ways as a quotient group of $G^{2,31,31}$. A similar remark holds for the duplication of the pair 33, 33.

QUARRY HEIGHTS, CANAL ZONE

* *Necessary and sufficient conditions*, p. 70.

† Journal of the London Mathematical Society, vol. 11 (1936), p. 106.

NOTE ON THE DEFINING RELATIONS FOR THE SIMPLE GROUP OF ORDER 660

F. A. LEWIS

Dickson* has given the following set of defining relations for the simple group of order 660:

$$(1) \qquad S^{11} = I, \quad T^2 = I, \quad (ST)^3 = I,$$

$$(2) \qquad (S^3TS^3T)^2 = I,$$

$$(3) \qquad (S^4TS^6T)^2 = I.$$

It follows from (3) that

$$TS^4T = S^5TS^7TS^5,$$

so that

$$\begin{aligned} TS^8T &= S^5TS^7TS^{10}TS^7TS^5 \\ &= S^5TS^8 \cdot S^{-1}TS^{-1}TS^{-1} \cdot S^8TS^5 \\ &= S^5TS^8TS^8TS^5. \end{aligned}$$

From this last relation

$$\begin{aligned} S^3TS^5TS^5 \cdot S^3TS^8T \cdot S^5T &= I, \\ S^3TS^8T &= S^6TS^6TS^8TS^6, \\ (S^3TS^8T)^2 &= S^6TS^6TS^8TSTS^6TS^8TS^6 \\ &= S^6TS^6TS^7 \cdot STSTS \cdot S^5TS^8TS^6 \\ &= S^6TS^6 \cdot TS^7TS^5T \cdot S^8TS^6 \\ &= S^6TS^6 \cdot S^6TS^4 \cdot S^8TS^6 \\ &= S^6TSTSTS^6 \\ &= I. \end{aligned}$$

Since (2) has been derived from (1) and (3), it follows that (1) and (3) are sufficient to define the simple group of order 660.

UNIVERSITY OF ALABAMA

* This Bulletin, vol. 9 (1903), pp. 204–206.